

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:19:41 ; Search time 17 Seconds
(without alignments)
980.144 Million cell updates/sec

Title: US-09-936-456-2

Perfect score: 320

Sequence: 1 MTLNMGVLPFYPPQPRHAGF.....GDPILHKQALPDLKCTITNL 320

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	8	2.5	615	1	MUTL_ECO57	Q8xdn4 escherichia
2	8	2.5	615	1	MUTL_ECO57	Q8fak9 escherichia
3	8	2.5	615	1	MUTL_ECOLI	P23367 escherichia
4	8	2.5	617	1	MUTL_PASMU	P57886 pasteurella
5	8	2.5	618	1	MUTL_SALTY	Q8z187 salmonella
6	8	2.5	618	1	MUTL_SALTY	P14161 salmonella
7	8	2.5	624	1	MUTL_CHUTE	Q8kax3 chlorobium
8	8	2.5	629	1	MUTL_HAETN	P44494 haemophilus
9	8	2.5	632	1	MUTL_PSEB	Q88dd1 pseudomonas
10	8	2.5	633	1	MUTL_PSEB	Q9hul8 pseudomonas
11	8	2.5	635	1	MUTL_YERPE	Q8ziw4 yersinia pe
12	8	2.5	645	1	MUTL_PSEB	Q87vj2 pseudomonas
13	8	2.5	645	1	MUTL_STABP	Q8cpe9 staphylococ
14	8	2.5	653	1	MUTL_VIBCH	Q9kvl3 vibrio chol
15	8	2.5	664	1	MUTL_VIBVU	Q8dcv0 vibrio vuln
16	8	2.5	669	1	MUTL_STAMU	Q99uh7 staphylococ
17	8	2.5	669	1	MUTL_STAMU	Q93t05 staphylococ
18	8	2.5	669	1	MUTL_STAMU	Q8hwx9 staphylococ
19	8	2.5	669	1	MUTL_VIBPA	Q87l05 vibrio para
20	8	2.5	49	1	COX4_PARDE	P77921 paracoccus
21	7	2.2	116	1	CHPB_ECOLI	P33647 escherichia
22	7	2.2	116	1	CN4D_MOUSE	Q01063 mus musculu
23	7	2.2	131	1	PRDC_PROVU	P20923 proteus vul
24	7	2.2	181	1	IGF2_PIG	P23695 sus scrofa
25	7	2.2	243	1	YCBR_BACSU	P42250 bacillus su
26	7	2.2	262	1	COBS_SHEON	Q8e117 shewanella
27	7	2.2	266	1	FLIV_ECOLI	Q39174 escherichia
28	7	2.2	277	1	ETB_STAUV	P09332 staphylococ
29	7	2.2	281	1	BRHB_COMTE	Q46381 c cis-2,3-d
30	7	2.2	288	1	YGHU_ECOLI	Q46845 escherichia
31	7	2.2	289	1	ATP6_PROMO	P19903 propionigen
32	7	2.2	299	1	T2R1_HUMAN	Q9nyw7 homo sapien
33	7	2.2	308	1	KHSE_CORGL	P07128 corynebacte

34	7	2.2	311	1	MRAM_STAAM	O07320 staphylococ
35	7	2.2	311	1	MRAM_STAEP	Q8csx7 staphylococ
36	7	2.2	313	1	ARAP_BACSU	P94529 bacillus su
37	7	2.2	315	1	KHSE_MYCLE	P45836 mycobacteri
38	7	2.2	316	1	KHSE_MYCTU	Q10603 mycobacteri
39	7	2.2	319	1	CCSA_OENHO	Q9mt12 oenothera h
40	7	2.2	324	1	YK95_MYCLE	P54075 mycobacteri
41	7	2.2	335	1	SM34_LYTP1	Q05904 lytechinus
42	7	2.2	344	1	MI2D_BACSU	P26935 bacillus su
43	7	2.2	345	1	DDL_WOLSU	Q7ma71 wolinnella s
44	7	2.2	401	1	ASSV_CHUTE	Q8kde0 chlorobium
45	7	2.2	418	1	NISC_IACIA	Q03202 lactococcus

ALIGNMENTS

RESULT 1
MUTL_ECO57
ID MUTL_ECO57 STANDARD; PRT; 615 AA.
AC Q8XDN4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA mismatch repair protein mult.
GN MUTL OR Z5777 OR ECS5146.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
Postal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
Weich R.A., Blattner F.R.;
RA "Genome sequence of enterohemorrhagic Escherichia coli O157:H7";
RT Nature 409:529-533(2001).
RL [2]
SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsuka E., Nakayama K., Murata T., Tanaka M., Tobe T.,
Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
Kunara S., Shiba T., Hattori M., Shiragawa H.;
RA "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12";
RT DNA Res. 8:11-22(2001).
RL
-!- FUNCTION: This protein is involved in the repair of mismatches in
DNA. It is required for dam-dependent methyl-directed DNA mismatch
repair. May act as a "molecular matchmaker", a protein that
promotes the formation of a stable complex between two or more
DNA-binding proteins in an ATP-dependent manner without itself
being part of a final effector complex (By similarity).
-!- SIMILARITY: Belongs to the DNA mismatch repair mult/hexB family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AE005650; AAG59366.1; -
DR EMBL; AP002568; BAB38569.1; -
DR PIR; B86113; B86113.
DR PIR; B91272; B91272.

DR HAMAP; MF_00149; -, 1.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR002099; DNA_mis_repair.
 DR Pfam; PF01119; DNA_mis_repair; 1.
 DR Pfam; PF02518; HATPase_c; 1.
 DR SMART; SM00387; HATPase_c; 1.
 DR TIGRFAMs; TIGR00585; mutL; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 DR DNA repair; Complete proteome.
 SQ SEQUENCE 615 AA; 67921 MW; 1B5F0F177AB11EB9 CRC64;

Query Match 2.5%; Score 8; DB 1; Length 615;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISV 230
 Db 98 ALASISV 105

RESULT 2

MUTL_ECOL6 STANDARD; PRT; 615 AA.

ID MUTL_ECOL6 STANDARD; PRT; 615 AA.
 AC O8FAK9;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE DNA mismatch repair protein mutL.
 GN MUTL OR C5254.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Lion S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 CC -!- FUNCTION: This protein is involved in the repair of mismatches in
 DNA. It is required for dam-dependent methyl-directed DNA mismatch
 repair. May act as a "molecular matchmaker", a protein that
 promotes the formation of a stable complex between two or more
 DNA-binding proteins in an ATP-dependent manner without itself
 being part of a final effector complex (By similarity).
 CC -!- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AE016771; AAN83676.1; ALT_INIT.
 DR HAMAP; MF_00149; -, 1.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR002099; DNA_mis_repair.
 DR Pfam; PF01119; DNA_mis_repair; 1.
 DR Pfam; PF02518; HATPase_c; 1.
 DR TIGRFAMs; TIGR00585; mutL; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 DR DNA repair; Complete proteome.
 KW SEQUENCE 615 AA; 67910 MW; 770FA42138081CBB CRC64;

Query Match 2.5%; Score 8; DB 1; Length 615;
 Best Local Similarity 100.0%; Pred. No. 12;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 223 ALASISV 230
 Db 98 ALASISV 105

RESULT 3

MUTL_ECOLI STANDARD; PRT; 615 AA.

ID MUTL_ECOLI STANDARD; PRT; 615 AA.
 AC P23367;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 26-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA mismatch repair protein mutL.
 GN MUTL OR B4170.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=92279041; PubMed=1594459;
 RA Tsui H.-C., Mandavilli B., Winkler M.E.;
 RT "Nonconserved segment of the MutL protein from Escherichia coli K-12
 and Salmonella typhimurium.";
 RL Nucleic Acids Res. 20:2379-2379(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=95334362; PubMed=7610040;
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
 RA Blattner F.R.;
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
 region from 92.8 through 100 minutes.";
 RL Nucleic Acids Res. 23:2105-2119(1995).
 RN [3]
 RP SEQUENCE OF 596-615 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=91154127; PubMed=1999389;
 RA Winkler M.E., Connolly D.M.;
 RT "Structure of Escherichia coli K-12 *miaA* and characterization of the
 mutator phenotype caused by *miaA* insertion mutations.";
 RL J. Bacteriol. 173:1711-1721(1991).
 RN [4]
 RP SEQUENCE OF 1-23 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=94195106; PubMed=7511774;
 RA Tsui H.-C., Zhao G., Feng G., Jeung H.-C., Winkler M.E.;
 RT "The mutL repair gene of Escherichia coli K-12 forms a superoperon
 with a gene encoding a new cell-wall amidase.";
 RL Mol. Microbiol. 11:189-202(1994).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 1-349.
 RX MEDLINE=99043508; PubMed=9827806;
 RA Ban C., Yang W.;
 RT "Crystal structure and ATPase activity of MutL: implications for DNA
 repair and mutagenesis.";
 RL Cell 95:541-552(1998).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-349.
 RC STRAIN=K12;
 RX MEDLINE=99213497; PubMed=10199405;
 RA Ban C., Junop M., Yang W.;
 RT "Transformation of MutL by ATP binding and hydrolysis: a switch in
 DNA mismatch repair.";
 RL Cell 97:85-97(1999).
 CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES IN
 CC DNA. IT IS REQUIRED FOR DAM-DEPENDENT METHYL-DIRECTED DNA MISMATCH
 CC REPAIR. MAY ACT AS A "MOLECULAR MATCHMAKER", A PROTEIN THAT
 CC PROMOTES THE FORMATION OF A STABLE COMPLEX BETWEEN TWO OR MORE
 CC DNA-BINDING PROTEINS IN AN ATP-DEPENDENT MANNER WITHOUT ITSELF

CC BEING PART OF A FINAL EFFECTOR COMPLEX. THE ATPASE ACTIVITY OF
 CC MUTL IS STIMULATED BY DNA.
 CC -I- SIMILARITY: Belongs to the DNA mismatch repair mutl/hexb family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z11831; CAA77850.1; -.
 DR EMBL; U14003; AAA97066.1; -.
 DR EMBL; AE000489; AAC77127.1; -.
 DR EMBL; M63655; AAA24173.1; -.
 DR EMBL; L19346; AAA20098.1; -.
 DR PIR; PH0853; PH0853.
 DR PDB; 1B62; 28-APR-99.
 DR PDB; 1B63; 08-JUN-99.
 DR PDB; 1BKN; 11-MAY-99.
 DR EcoGene; EG11281; mutL.
 DR HAMAP; MF_00149; -; 1.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR002099; DNA_mis_repair.
 DR Pfam; PF01119; DNA_mis_repair; 1.
 DR Pfam; PF02518; HATPase_c; 1.
 DR SMART; SM00387; HATPase_c; 1.
 DR TIGRfams; TIGR00585; mutl; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 DR KW DNA repair; ATP-binding; 3D-structure; Complete proteome.
 FT HELIX 8 19
 FT HELIX 23 36
 FT TURN 37 38
 FT STRAND 41 47
 FT HELIX 48 50
 FT TURN 51 51
 FT STRAND 53 58
 FT HELIX 65 67
 FT HELIX 68 72
 FT TURN 74 75
 FT HELIX 83 87
 FT TURN 88 88
 FT HELIX 98 103
 FT TURN 104 105
 FT STRAND 107 113
 FT TURN 115 116
 FT STRAND 120 126
 FT TURN 127 130
 FT STRAND 131 137
 FT STRAND 142 148
 FT TURN 150 153
 FT HELIX 155 159
 FT TURN 160 160
 FT HELIX 164 181
 FT TURN 183 184
 FT STRAND 186 191
 FT TURN 192 193
 FT STRAND 194 199
 FT TURN 204 205
 FT TURN 208 208
 FT HELIX 209 216
 FT HELIX 218 223
 FT STRAND 224 231
 FT TURN 232 233
 FT STRAND 234 241
 FT HELIX 243 245
 FT TURN 248 249
 FT HELIX 250 252
 FT STRAND 254 258
 FT TURN 259 260
 FT STRAND 261 262
 FT HELIX 266 279

FT STRAND 288 293
 FT HELIX 296 298
 FT STRAND 299 300
 FT TURN 302 303
 FT TURN 305 306
 FT STRAND 311 312
 FT TURN 313 314
 FT HELIX 315 329
 FT TURN 330 330
 SQ SEQUENCE 615 AA; 67923 MW; C69D2735BF5FA165 CRC64;

Query Match 2.5%; Score 8; DB 1; Length 615;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALA5ISSV 230
 Db 98 ALA5ISSV 105

RESULT 4
 MUTL_PASMU
 ID MUTL_PASMU STANDARD; PRT; 617 AA.

AC P57886;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA mismatch repair protein mutL.

GN MUTL OR PM0904.

OS Pasteurella multocida.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Pasteurella.

OX NCBI_TaxID=747;

RN [1]

SEQUENCE FROM N.A.

RP STRAIN=Pm70;

RX MEDLINE=21145866; PubMed=11248100;

RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;

RT "Complete genomic sequence of Pasteurella multocida Pm70.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).

CC -I- FUNCTION: This protein is involved in the repair of mismatches in

DNA. It is required for dam-dependent methyl-directed DNA mismatch

repair. May act as a "molecular matchmaker", a protein that

promotes the formation of a stable complex between two or more

DNA-binding proteins in an ATP-dependent manner without itself

being part of a final effector complex (By similarity).

CC -I- SIMILARITY: Belongs to the DNA mismatch repair mutl/hexb family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AE006129; AAK02988.1; -.

DR HSSP; P23367; 1BKN.

DR HAMAP; MF_00149; -; 1.

DR InterPro; IPR003594; ATPbind_ATPase.

DR InterPro; IPR002099; DNA_mis_repair.

DR Pfam; PF01119; DNA_mis_repair; 1.

DR Pfam; PF02518; HATPase_c; 1.

DR SMART; SM00387; HATPase_c; 1.

DR TIGRfams; TIGR00585; mutl; 1.

DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.

DR KW DNA repair; Complete proteome.

SQ SEQUENCE 617 AA; 69798 MW; 8DC85F15695460AE CRC64;

Query Match 2.5%; Score 8; DB 1; Length 617;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 223 ALASISSV 230
 Db 98 ALASISSV 105

RESULT 5

MUTL_SALT
 ID MUTL_SALT STANDARD; PRT: 618 AA.
 AC Q8Z187;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE DNA mismatch repair protein mutL.
 GN MUTL OR STY4716 OR T4410.
 OS Salmonella typhi.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagsels K.,
 RA Krogan A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrell B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).

CC -1- FUNCTION: This protein is involved in the repair of mismatches in
 CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
 CC repair. May act as a "molecular matchmaker", a protein that
 CC promotes the formation of a stable complex between two or more
 CC DNA-binding proteins in an ATP-dependent manner without itself
 CC being part of a final effector complex (By similarity).
 CC -1- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AL627283; CAD06836.1; -.
 DR EMBL; AE016849; AA071859.1; -.
 DR HAMAP; MF_00149; -; 1.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR002099; DNA_mis_repair.
 DR Pfam; PF01119; DNA_mis_repair; 1.
 DR SMART; SM00387; HATPase_c; 1.
 DR TIGRFAMs; TIGR00585; mutL; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 KW DNA repair; Complete proteome.
 SQ SEQUENCE 618 AA; 67829 MW; 18DD90043D5180F4 CRC64;

Query Match 2.5%; Score 8; DB 1; Length 618;
 Best local Similarity 100.0%; Pred. No. 12;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 223 ALASISSV 230
 Db 98 ALASISSV 105

RESULT 6

MUTL_SALT
 ID MUTL_SALT STANDARD; PRT: 618 AA.
 AC P14161;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA mismatch repair protein mutL.
 GN MUTL OR STM4359.
 OS Salmonella typhimurium.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=LT2;
 RX MEDLINE=90008766; PubMed=2676972;
 RA Mankovich J.A., McIntyre C.A., Walker G.C.;
 RT "Nucleotide sequence of the Salmonella typhimurium mutL gene required
 RT for mismatch repair: homology of MutL to HexB of Streptococcus
 RT pneumoniae and to PMS1 of the yeast Saccharomyces cerevisiae.";
 RL J. Bacteriol. 171:5325-5331(1989).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).

CC -1- FUNCTION: This protein is involved in the repair of mismatches in
 CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
 CC repair. May act as a "molecular matchmaker", a protein that
 CC promotes the formation of a stable complex between two or more
 CC DNA-binding proteins in an ATP-dependent manner without itself
 CC being part of a final effector complex.
 CC -1- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; M29687; AAA27166.1; -.
 DR EMBL; AE008904; AAL23179.1; -.
 DR PIR; A33588; A33588.
 DR HSSP; P23367; 1B62.
 DR StGene; SG10236; mutL.
 DR HAMAP; MF_00149; -; 1.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR002099; DNA_mis_repair.
 DR Pfam; PF01119; DNA_mis_repair; 1.
 DR SMART; SM00387; HATPase_c; 1.
 DR TIGRFAMs; TIGR00585; mutL; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 KW DNA repair; Complete proteome.
 SQ SEQUENCE 618 AA; 67762 MW; 2896CA26AD358584 CRC64;

Query Match 2.5%; Score 8; DB 1; Length 618;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
 |||||
 Db 98 ALASISSV 105

RESULT 7

MUTL_CHLTB
 ID MUTL_CHLTB STANDARD; PRT; 624 AA.
 AC Q8KAX3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA mismatch repair protein mutL.
 GN MUTL OR CT2028.
 OS Chlorobium tepidum.
 OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
 OC Chlorobium.
 OX NCBI_TaxID=1097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TLS / ATCC 49652 / DSM 12025;
 RX MEDLINE=22103685; PubMed=12093901;
 RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
 RA Dodson R.J., Deboy R., Gwin M.L., Nelson W.C., Haft D.H.,
 RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
 RA Holt I., Mayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
 RA Niernan W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
 RA Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
 RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
 RT "The complete genome sequence of Chlorobium tepidum TLS, a
 photosynthetic, anaerobic, green-sulfur bacterium.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
 CC -!- FUNCTION: This protein is involved in the repair of mismatches in
 DNA. It is required for dam-dependent methyl-directed DNA mismatch
 repair. May act as a "molecular matchmaker", a protein that
 promotes the formation of a stable complex between two or more
 DNA-binding proteins in an ATP-dependent manner without itself
 being part of a final effector complex (By similarity).
 CC -!- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB012953; AAW73245.1; -.
 DR TIGR; CT2028; -.
 DR HAMAP; MF_00149; -; 1.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR002099; DNA_mis_repair.
 DR Pfam; PF01119; DNA_mis_repair; 1.
 DR Pfam; PF02518; HATPase_c; 1.
 DR TIGRFAMs; TIGR00585; mutL; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 KM DNA repair; Complete proteome.
 SQ SEQUENCE 624 AA; 69836 MW; 43F3CE8234A42CFF CRC64;

Query Match 2.5%; Score 8; DB 1; Length 624;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
 |||||
 Db 99 ALASISSV 106

RESULT 8

MUTL_HABIN
 ID MUTL_HABIN STANDARD; PRT; 629 AA.
 AC P44494;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA mismatch repair protein mutL.
 GN MUTL OR HI0067.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kierlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA Mckenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Sadek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Furmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd.";
 RL Science 269:496-512(1995).
 CC -!- FUNCTION: This protein is involved in the repair of mismatches in
 DNA. It is required for dam-dependent methyl-directed DNA mismatch
 repair. May act as a "molecular matchmaker", a protein that
 promotes the formation of a stable complex between two or more
 DNA-binding proteins in an ATP-dependent manner without itself
 being part of a final effector complex (By similarity).
 CC -!- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U32692; AAC21745.1; -.
 DR PIR; E64046; E64046.
 DR HSSP; P23367; 1BKN.
 DR TIGR; HI0067; -.
 DR HAMAP; MF_00149; -; 1.
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR InterPro; IPR002099; DNA_mis_repair.
 DR Pfam; PF01119; DNA_mis_repair; 1.
 DR Pfam; PF02518; HATPase_c; 1.
 DR SMART; SM00387; HATPase_c; 1.
 DR TIGRFAMs; TIGR00585; mutL; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 KM DNA repair; Complete proteome.
 SQ SEQUENCE 629 AA; 71622 MW; 0ADC240FD94D1556 CRC64;

Query Match 2.5%; Score 8; DB 1; Length 629;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
 |||||
 Db 98 ALASISSV 105

RESULT 9
 MUTL_PSEPK
 ID MUTL_PSEPK STANDARD; PRT; 632 AA.
 AC Q88D11;

```

DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DNA mismatch repair protein mutL.
GN MUTL OR PA4896.
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzez A.,
RA Usterback T., Rizzo M., Lee K., Kosack D., Moestl D., Medler H.,
RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA Kiewitz C., Eisen J.A., Timmis K.N., Duesterhoeft A., Tuenmler B.,
RA Fraser C.M.;
RA "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440.";
RL Environ. Microbiol. 4:799-808(2002).
CC -1- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
CC repair. May act as a "molecular matchmaker", a protein that
CC promotes the formation of a stable complex between two or more
CC DNA-binding proteins in an ATP-dependent manner without itself
CC being part of a final effector complex (By similarity).
CC -1- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE016792; AAN70463.1; -.
DR TIGR; PA4896; -.
DR HAMAP; MF_00149; -.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR002099; DNA_mis_repair.
DR Pfam; PF01119; DNA_mis_repair; 1.
DR Pfam; PF02518; HATPase_C; 1.
DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
KW DNA repair; Complete proteome.
SQ SEQUENCE 632 AA; 69745 MW; F7B93ECF5F78AD8F CRC64;

Query Match 2.5%; Score 8; DB 1; Length 632;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALA5ISSV 230
DB 102 ALA5ISSV 109

RESULT 10
MUTL_PSEAE STANDARD; PRT; 633 AA.
ID MUTL_PSEAE
AC Q9HUI8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA mismatch repair protein mutL.
GN MUTL OR PA4946.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Jory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
CC repair. May act as a "molecular matchmaker", a protein that
CC promotes the formation of a stable complex between two or more
CC DNA-binding proteins in an ATP-dependent manner without itself
CC being part of a final effector complex (By similarity).
CC -1- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE004907; AAG08331.1; -.
DR PIR; F83028; F83028.
DR HSSP; P23367; 1BKN.
DR HAMAP; MF_00149; -.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR002099; DNA_mis_repair.
DR Pfam; PF01119; DNA_mis_repair; 1.
DR Pfam; PF02518; HATPase_C; 1.
DR SMART; SM00387; HATPase_C; 1.
DR TIGRFAMS; TIGR00585; mutL; 1.
DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
KW DNA repair; Complete proteome.
SQ SEQUENCE 633 AA; 69747 MW; 3FD58AF0A721FE2D CRC64;

Query Match 2.5%; Score 8; DB 1; Length 633;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALA5ISSV 230
DB 102 ALA5ISSV 109

RESULT 11
MUTL_YERP3 STANDARD; PRT; 635 AA.
ID MUTL_YERP3
AC Q8ZIW4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA mismatch repair protein mutL.
GN MUTL OR YPO0371 OR Y0628.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,

```

```
RA Feltwell T., Hamlin N., Holroyd S., Tagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.,
RT "Genome sequence of Yersinia pestis, the causative agent of plague.",
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5 / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).
CC -!- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
CC repair. May act as a "molecular matchmaker", a protein that
CC promotes the formation of a stable complex between two or more
CC DNA-binding proteins in an ATP-dependent manner without itself
CC being part of a final effector complex (By similarity).
CC -!- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ414142; CAC89230.1; -.
DR EMBL; AE013665; AAM84216.1; -.
DR PIR; AC0046; AC0046.
DR HAMAP; MF_00149; -.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR002099; DNA_mis_repair.
DR Pfam; PF01119; DNA_mis_repair; 1.
DR Pfam; PF02518; HATPase_c; 1.
DR SMART; SM00387; HATPase_c; 1.
DR TIGRFAMs; TIGR00585; mutL; 1.
DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
DR DNA repair; Complete proteome.
SQ SEQUENCE 635 AA; 70287 MW; 301CA81150449D45 CRC64;

Query Match 2.5%; Score 8; DB 1; Length 635;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 98 ALASISSV 105

RESULT 12
MUTL_PSESMS STANDARD; PRT; 645 AA.
ID MUTL_PSESMS
AC Q87VJ2;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DNA mismatch repair protein mutL.
GN MUTL OR PSPTO4944.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=22834015; PubMed=12928499;
RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
```

```
RA Gwin M.L., Dodson R.J., Deboy R.T., Durkin A.S., Kolonay J.F.,
RA Madupu R., Dagherty S., Brinkac L., Beanan M.J., Haft D.H.,
RA Nelson W.C., Davidsen T., Zafar N., Zhou L., Liu J., Yuan Q.,
RA Khouri H., Fedorova N., Tran B., Russell D., Berry K., Uitterback T.,
RA Van Aken S.E., Feldblyum T.V., D'Ascenzo M., Deng W.-L., Ramos A.R.,
RA Alfano J.R., Cartinhour S., Chatterjee A.K., Delaney T.P.,
RA Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X., Bender C.L.,
RA White O., Fraser C.M., Collier A.;
RT "The complete genome sequence of the Arabidopsis and tomato pathogen
RT Pseudomonas syringae pv. tomato DC3000.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
CC -!- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
CC repair. May act as a "molecular matchmaker", a protein that
CC promotes the formation of a stable complex between two or more
CC DNA-binding proteins in an ATP-dependent manner without itself
CC being part of a final effector complex (By similarity).
CC -!- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE016873; AA058372.1; -.
DR TIGR; PSPTO4944; -.
DR HAMAP; MF_00149; -.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR002099; DNA_mis_repair.
DR Pfam; PF01119; DNA_mis_repair; 1.
DR Pfam; PF02518; HATPase_c; 1.
DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
DR DNA repair; Complete proteome.
SQ SEQUENCE 645 AA; 70972 MW; 58245DE1BF44515F CRC64;

Query Match 2.5%; Score 8; DB 1; Length 645;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 118 ALASISSV 125

RESULT 13
MUTL_STAEP STANDARD; PRT; 645 AA.
ID MUTL_STAEP
AC Q8CPB9;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DNA mismatch repair protein mutL.
GN MUTL OR SE0975.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RX PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593(2003).
CC -!- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is required for dam-dependent methyl-directed DNA mismatch
CC repair. May act as a "molecular matchmaker", a protein that
CC promotes the formation of a stable complex between two or more
```

```
CC      DNA-binding proteins in an ATP-dependent manner without itself
CC      being part of a final effector complex (By similarity).
CC      -!- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
CC
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE016747; AAC04572.1; -.
DR      HAMAP; MF_00149; -.
DR      InterPro; IPR003594; ATPbind_ATPase.
DR      InterPro; IPR002099; DNA_mis_repair.
DR      Pfam; PF01119; DNA_mis_repair; 1.
DR      Pfam; PF02518; HATPase_c; 1.
DR      TIGRFAMs; TIGR00585; mutL; 1.
DR      PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
KW      DNA repair; Complete proteome.
SQ      SEQUENCE 645 AA; 73526 MW; FFA17BD6C3E47DC9 CRC64;

Query Match      2.5%; Score 8; DB 1; Length 645;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      223 ALASISSV 230
Db      99 ALASISSV 106

RESULT 14
MUTL_VIBCH
ID      MUTL_VIBCH      STANDARD;      PRT;      653 AA.
AC      Q9KVL3;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      DNA mismatch repair protein mutL.
GN      MUTL OR VC0345.
OS      Vibrio cholerae.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OX      Vibrionaceae; Vibrio.
NX      NCBI_TaxID=666;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=El Tor N16961 / Serotype O1;
RX      MEDLINE=20406833; PubMed=10952301;
RA      Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA      Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA      Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA      Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA      McDonald L., Utterback T., Fleischmann R.D., Nieman W.C., White O.,
RA      Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA      Fraser C.M.;
RT      "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT      cholerae.";
RL      Nature 406:477-483(2000).
CC      -!- FUNCTION: This protein is involved in the repair of mismatches in
CC      DNA. It is required for dam-dependent methyl-directed DNA mismatch
CC      repair. May act as a "molecular matchmaker", a protein that
CC      promotes the formation of a stable complex between two or more
CC      DNA-binding proteins in an ATP-dependent manner without itself
CC      being part of a final effector complex (By similarity).
CC      -!- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE004123; AAF93518.1; -.
DR      PIR; A82334; A82334.
DR      HSSP; P23367; IBRN.
DR      TIGR; VC0345; -.
DR      HAMAP; MF_00149; -.
DR      InterPro; IPR003594; ATPbind_ATPase.
DR      InterPro; IPR002099; DNA_mis_repair.
DR      Pfam; PF01119; DNA_mis_repair; 1.
DR      Pfam; PF02518; HATPase_c; 1.
DR      SMART; SM00387; HATPase_c; 1.
DR      TIGRFAMs; TIGR00585; mutL; 1.
DR      PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
KW      DNA repair; Complete proteome.
SQ      SEQUENCE 653 AA; 72630 MW; 5D970C21A17666FF CRC64;

Query Match      2.5%; Score 8; DB 1; Length 653;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      223 ALASISSV 230
Db      98 ALASISSV 105

RESULT 15
MUTL_VIBVU
ID      MUTL_VIBVU      STANDARD;      PRT;      664 AA.
AC      Q8DCV0;
DT      10-OCT-2003 (Rel. 42, Created)
DT      10-OCT-2003 (Rel. 42, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      DNA mismatch repair protein mutL.
GN      MUTL OR VV11292.
OS      Vibrio vulnificus.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OX      Vibrionaceae; Vibrio.
NX      NCBI_TaxID=672;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CMCP6;
RA      Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA      Choy H.E.;
RT      "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL      Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
CC      -!- FUNCTION: This protein is involved in the repair of mismatches in
CC      DNA. It is required for dam-dependent methyl-directed DNA mismatch
CC      repair. May act as a "molecular matchmaker", a protein that
CC      promotes the formation of a stable complex between two or more
CC      DNA-binding proteins in an ATP-dependent manner without itself
CC      being part of a final effector complex (By similarity).
CC      -!- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE016801; AAC09747.1; -.
DR      HAMAP; MF_00149; -.
DR      InterPro; IPR003594; ATPbind_ATPase.
DR      InterPro; IPR002099; DNA_mis_repair.
DR      Pfam; PF01119; DNA_mis_repair; 1.
DR      Pfam; PF02518; HATPase_c; 1.
DR      TIGRFAMs; TIGR00585; mutL; 1.
DR      PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
KW      DNA repair; Complete proteome.
SQ      SEQUENCE 664 AA; 74006 MW; A0ECF5EF88247A7C CRC64;
```


Query Match 2.5%; Score 8; DB 1; Length 664;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 98 ALASISSV 105

Search completed: February 23, 2004, 19:33:49
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:19:21 ; Search time 59 Seconds
(without alignments)
1532.461 Million cell updates/sec

Title: US-09-936-456-2

Perfect score: 320

Sequence: 1 MTLNMGVLPFYQPRHAAGF.....GDP LHKQALPDLKITTNL 320

Scoring table: OLIGO

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	320	100.0	320	3 AAB18992	Aab18992 Amino aci
2	320	100.0	320	3 AAB26325	Aab26325 Human CAS
3	220	68.8	242	3 AAB52095	Aab52095 Human sec
4	219	68.4	320	5 ABB90126	Abb90126 Human pol
5	219	68.4	370	7 ADD19303	Add19303 Human sec
6	172	53.8	262	3 AAB26326	Aab26326 CASB618 p
7	49	15.3	103	7 ADD19262	Add19262 Human col
8	19	5.9	98	4 AAG76780	Aag76780 Human sec
9	19	5.9	112	4 AAB66073	Aab66073 Human INT
10	19	5.9	112	6 ABO32727	AbO32727 Secreted
11	19	5.9	298	6 ADA55266	Ada55266 Human pro
12	19	5.9	343	4 AAB66071	Aab66071 Human INT
13	19	5.9	343	6 ABU08370	Abu08370 Amino aci
14	19	5.9	343	6 ABO32725	AbO32725 Secreted
15	19	5.9	445	4 ABO33520	AbO33520 Novel hum
16	13	4.1	343	5 ABB97510	Abb97510 Novel hum
17	10	3.1	10	3 AAB26378	Aab26378 Human CAS
18	10	3.1	10	3 AAB26397	Aab26397 Human CAS
19	10	3.1	10	3 AAB26399	Aab26399 Human CAS
20	10	3.1	10	3 AAB26371	Aab26371 Human CAS
21	10	3.1	10	3 AAB26381	Aab26381 Human CAS
22	10	3.1	10	3 AAB26396	Aab26396 Human CAS
23	10	3.1	10	3 AAB26369	Aab26369 Human CAS
24	10	3.1	10	3 AAB26373	Aab26373 Human CAS
25	10	3.1	10	3 AAB26382	Aab26382 Human CAS

26	10	3.1	10	3 AAB26379	Aab26379 Human CAS
27	10	3.1	10	3 AAB26392	Aab26392 Human CAS
28	10	3.1	10	3 AAB26375	Aab26375 Human CAS
29	10	3.1	10	3 AAB26384	Aab26384 Human CAS
30	10	3.1	10	3 AAB26383	Aab26383 Human CAS
31	10	3.1	10	3 AAB26389	Aab26389 Human CAS
32	10	3.1	10	3 AAB26370	Aab26370 Human CAS
33	10	3.1	10	3 AAB26372	Aab26372 Human CAS
34	10	3.1	10	3 AAB26376	Aab26376 Human CAS
35	10	3.1	10	3 AAB26387	Aab26387 Human CAS
36	10	3.1	10	3 AAB26395	Aab26395 Human CAS
37	10	3.1	10	3 AAB26385	Aab26385 Human CAS
38	10	3.1	10	3 AAB26374	Aab26374 Human CAS
39	10	3.1	10	3 AAB26393	Aab26393 Human CAS
40	10	3.1	10	3 AAB26386	Aab26386 Human CAS
41	10	3.1	10	3 AAB26398	Aab26398 Human CAS
42	10	3.1	10	3 AAB26377	Aab26377 Human CAS
43	10	3.1	10	3 AAB26380	Aab26380 Human CAS
44	10	3.1	10	3 AAB26390	Aab26390 Human CAS
45	10	3.1	10	3 AAB26391	Aab26391 Human CAS

ALIGNMENTS

RESULT 1	
AAB18992	
ID	AAB18992 standard; protein; 320 AA.
XX	
AC	AAB18992;
XX	
DT	08-FEB-2001 (first entry)
XX	
DE	Amino acid sequence of a human transmembrane protein.
XX	
KW	Human; transmembrane protein; cell proliferation disorder; myeloma;
KW	reproductive disorder; smooth muscle disorder; neurological disorder;
KW	arteriosclerosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
KW	allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;
KW	Alzheimer's disease; Tourette's disorder.
XX	
OS	Homo sapiens.
XX	
PH	
FT	Key
FT	Modified-site
FT	Location/Qualifiers
FT	84
FT	/note= "potential glycosylation site"
FT	86
FT	/note= "potential phosphorylation site"
FT	96
FT	/note= "potential phosphorylation site"
FT	109
FT	/note= "potential glycosylation site"
FT	121
FT	/note= "potential glycosylation site"
FT	131
FT	/note= "potential glycosylation site"
FT	156
FT	/note= "potential glycosylation site"
FT	280
FT	/note= "potential phosphorylation site"
FT	285
FT	/note= "potential phosphorylation site"
XX	
PN	WO200056891-A2.
XX	
PD	28-SEP-2000.
XX	
PF	22-MAR-2000; 2000WO-US007817.
XX	
PR	22-MAR-1999; 99US-0125537P.
PR	16-JUN-1999; 99US-0139565P.
XX	
PA	(INCY-) INCYTE PHARM INC.

```

XX Yue H, Lal P, Tang YT, Hillman JL, Reddy R, Bandman O;
PI Baughn MR, Lu DAM, Azimzai Y, Yang J;
XX WPI; 2000-579485/54.
DR N-PSDB; AAA96505.
XX
PT New human transmembrane proteins are used to treat a disease or condition
PT associated with decreased expression of functional HTMP e.g. Tourette's
PT disorder, angina and leukemia.
XX
PS Claim 1; Page 109; 130pp; English.
XX
CC The present sequence represents a human transmembrane proteins (HTMP).
CC Agonists and antagonists of the protein are used to treat a disease or
CC condition associated with overexpression of the protein. Diseases and
CC conditions which can be treated include cell proliferative,
CC immunological, reproductive, smooth muscle and neurological disorders
CC e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency
CC syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,
CC stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The
CC polynucleotides may be used to detect and quantify gene expression in
CC biopsied tissues where protein expression may be correlated with disease
CC e.g. to determine absence, presence or excess expression of HTMP or to
CC monitor regulation of HTMP expression during therapeutic intervention
XX
SQ Sequence 320 AA;

Query Match      100.0%; Score 320; DB 3; Length 320;
Best Local Similarity 100.0%; Pred. No. 1.2e-300;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTLWNGVLPFYPPQPRHAAGSVPLLIIVLFLALAAFFLLIPGIRGHSRWFVLVRLLS 60
DB 1 MTLWNGVLPFYPPQPRHAAGSVPLLIIVLFLALAAFFLLIPGIRGHSRWFVLVRLLS 60
QY 61 LFIGAEIIVAHFSAEMFVGTNTNTSYKAFSAARVTARVGLLVGEGINITLTGTPVHQL 120
DB 61 LFIGAEIIVAHFSAEMFVGTNTNTSYKAFSAARVTARVGLLVGEGINITLTGTPVHQL 120
QY 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLYHOYHLAGHYA 180
DB 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLYHOYHLAGHYA 180
QY 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLYHOYHLAGHYA 180
DB 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLYHOYHLAGHYA 180
QY 181 SATLWVAFCEFWLISNVLSTPAPLYGGLALLTTGAFALFGVFALASISSVPLCPRLGSS 240
DB 181 SATLWVAFCEFWLISNVLSTPAPLYGGLALLTTGAFALFGVFALASISSVPLCPRLGSS 240
QY 241 ALTTQYGAAFWVTLATGVLCFLGGAVVSLOVVRPSALRTLDDQAKDCSQERGSSPLIL 300
DB 241 ALTTQYGAAFWVTLATGVLCFLGGAVVSLOVVRPSALRTLDDQAKDCSQERGSSPLIL 300
QY 301 GDP LHKQALPDLKCITTNL 320
DB 301 GDP LHKQALPDLKCITTNL 320

RESULT 2
AAB26325
ID AAB26325 standard; protein; 320 AA.
XX
AC AAB26325;
XX
DT 11-JAN-2001 (first entry)
XX
DE Human CASB618 protein.
XX
KW Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian;
KW colon; autoimmune disease; HLA_A0201.
XX
OS Homo sapiens.
XX
PN WO200053748-A2.

```

```

XX 14-SEP-2000.
XX
XX 09-MAR-2000; 2000MO-EP002048.
XX
XX 11-MAR-1999; 99GB-00005607.
XX 01-SEP-1999; 99GB-00020590.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Bruck CEM, Cassart J, Coche T, Vinals Y De Bassollec;
XX
DR WPI; 2000-572268/53.
DR N-PSDB; AAA94623.
XX
PT New human CASB618 polypeptide, useful as a vaccine for prophylactic and
PT therapeutic treatment of cancers, particularly ovarian or colon cancer,
PT autoimmune diseases and related conditions.
XX
PS Claim 1; Page 61-62; 76pp; English.
XX
CC The present sequence is human CASB618 protein. The gene for human CASB618
CC is thought to be located on chromosome 15. The present protein and
CC epitopes of this protein (see AAB26327 to AAB26399) are useful in
CC diagnosing the occurrence of tumour cells and in vaccines for prophylactic
CC and therapeutic treatment of cancers, particularly ovarian or colon
CC cancer, autoimmune diseases and related conditions
XX
SQ Sequence 320 AA;

Query Match      100.0%; Score 320; DB 3; Length 320;
Best Local Similarity 100.0%; Pred. No. 1.2e-300;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTLWNGVLPFYPPQPRHAAGSVPLLIIVLFLALAAFFLLIPGIRGHSRWFVLVRLLS 60
DB 1 MTLWNGVLPFYPPQPRHAAGSVPLLIIVLFLALAAFFLLIPGIRGHSRWFVLVRLLS 60
QY 61 LFIGAEIIVAHFSAEMFVGTNTNTSYKAFSAARVTARVGLLVGEGINITLTGTPVHQL 120
DB 61 LFIGAEIIVAHFSAEMFVGTNTNTSYKAFSAARVTARVGLLVGEGINITLTGTPVHQL 120
QY 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLYHOYHLAGHYA 180
DB 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLYHOYHLAGHYA 180
QY 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLYHOYHLAGHYA 180
DB 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLYHOYHLAGHYA 180
QY 181 SATLWVAFCEFWLISNVLSTPAPLYGGLALLTTGAFALFGVFALASISSVPLCPRLGSS 240
DB 181 SATLWVAFCEFWLISNVLSTPAPLYGGLALLTTGAFALFGVFALASISSVPLCPRLGSS 240
QY 241 ALTTQYGAAFWVTLATGVLCFLGGAVVSLOVVRPSALRTLDDQAKDCSQERGSSPLIL 300
DB 241 ALTTQYGAAFWVTLATGVLCFLGGAVVSLOVVRPSALRTLDDQAKDCSQERGSSPLIL 300
QY 301 GDP LHKQALPDLKCITTNL 320
DB 301 GDP LHKQALPDLKCITTNL 320

RESULT 3
AAB52095
ID AAB52095 standard; protein; 242 AA.
XX
AC AAB52095;
XX
DT 21-FEB-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 44 SEQ ID NO:144.
XX
KW Human, secreted protein; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antitumor; anticonvulsant; antibacterial;
KW antifungal; antiparasitic; cardiant; cancer; immune disease; allergy;

```

KW cardiovascular disorder; wound healing; infection; neurological disease.
XX
OS Homo sapiens.
XX
PN WO200061596-A1.
XX
PD 19-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US008983.
XX
PR 09-APR-1999; 99US-0128703P.
XX PR 14-JAN-2000; 2000US-0176068P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Ruben SM, Komatsoulis G;
XX
DR MPI; 2000-611865/58.
DR N-PSDB; AAC95564.
XX
XX
PT Fifty nucleic acid molecules encoding human secreted proteins, useful in
PT the prevention, treatment and diagnosis of cancer, immune disorders,
PT cardiovascular disorders and neurological diseases.
XX
PS Disclosure; Page 75-76; 505pp; English.
XX
CC Polynucleotide sequences AAC95521 - AAC95570 represent cDNA encoding 50
CC human secreted proteins AAB52012 - AAB52161. Sequences AAB52062 -
CC AAB52103 represent alternative polypeptides encoded by the genes, and
CC amino acid sequences with which they share homology. The genes and
CC proteins have activities dependent on the tissues and cells in which they
CC are expressed. Examples of their activities include cytostatic;
CC immunosuppressive; nootropic; neuroprotective; antiviral; antiallergic;
CC hepatotropic; antidiabetic; antiinflammatory; antitumor; antiparasitic;
CC anticonvulsant; antibacterial; antifungal; antiparasitic; and cardiant.
CC The secreted proteins, polypeptides, antagonists and agonists may be
CC useful in treating, preventing and/or diagnosing diseases and disorders
CC such as cancer, particularly breast and ovarian cancer, and other cancers
CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,
CC liver, lung, or urogenital. Immune disorders such as Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemia; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases such as viral,
CC bacterial, fungal and parasitic infections may also be treated using the
CC proteins and polynucleotides of the invention. Sequences AAC95512 -
CC AAC95520 and AAB52011 are used in the isolation and characterisation of
CC the proteins and polynucleotides of the invention
XX
SQ Sequence 242 AA;
Query Match 68.8%; Score 220; DB 3; Length 242;
Best Local Similarity 100.0%; Pred. No. 4.9e-204;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 44 GIRGSRWFVILVRLSLFVIGABIVAVHFSAEWFGVTNTNTSYKAFSARVTARVGLLV 103
Db 1 GIRGSRWFVILVRLSLFVIGABIVAVHFSAEWFGVTNTNTSYKAFSARVTARVGLLV 60
QY 104 GLEGINITLTGTPVHQINETIDYNEQFTWRLKENYAAYANALEKGLDPVLYIAEKPTP 163
Db 61 GLEGINITLTGTPVHQINETIDYNEQFTWRLKENYAAYANALEKGLDPVLYIAEKPTP 120
QY 164 SSPGGLYHQYHLAGHYASATLWVAFCFWLLSNVLLSTPAPLYGGLALTGTGAFALFGVFA 223
Db 121 SSPGGLYHQYHLAGHYASATLWVAFCFWLLSNVLLSTPAPLYGGLALTGTGAFALFGVFA 180
QY 224 LASISSVPLCPRLIGSSALLTQYGAFFWVTLATGVLCLFL 263
Db 181 LASISSVPLCPRLIGSSALLTQYGAFFWVTLATGVLCLFL 220

RESULT 4
AAB90126
ID ABB90126 standard; protein; 320 AA.
XX
AC ABB90126;
XX
DT 24-MAY-2002 (first entry)
XX
DE Human polypeptide SEQ ID NO 2502.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX
PN WO200190304-A2.
XX
PD 29-NOV-2001.
XX
PF 18-MAY-2001; 2001WO-US016450.
XX
PR 19-MAY-2000; 2000US-0205515P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Birse CE, Rosen CA;
XX
DR MPI; 2002-122018/16.
DR N-PSDB; ABL90535.
XX
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders.
XX
PS Claim 11; SEQ ID NO 2502; 2081pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABU89449-ABU90853) and proteins
CC (AAB89040-AAB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 320 AA;
Query Match 68.4%; Score 219; DB 5; Length 320;
Best Local Similarity 99.7%; Pred. No. 6e-203;
Matches 319; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MTLWNGVLPFYPPRHAAGFSVPLILVLFALAAFLILFGIRGSRWLVRLVLS 60
Db 1 MTLWNGVLPFYPPRHAAGFSVPLILVLFALAAFLILFGIRGSRWLVRLVLS 60
QY 61 LFIGABIVAVHFSAEWFGVTNTNTSYKAFSARVTARVGLVGLGINTITLTGTPVHQ 120
Db 61 LFIGABIVAVHFSAEWFGVTNTNTSYKAFSARVTARVGLVGLGINTITLTGTPVHQ 120

QY	121	NETIDYNEQFTWRLKENYAAEYANALEKGLPDPVLYLAEKFTPSSPCGLYHQYHLAGHYA	180
Db	121	NETIDYNEQFTWRLKENYAAEYANALEKGLPDPVLYLAEKFTPSSPCGLYHQYHLAGHYA	180
QY	181	SATLWVAFCFWLLSNVLLSTPAPLYGGLALLTTGAFAFLFGVAFALASISSVPLCPLRLGSS	240
Db	181	SATLWVAFCFWLLSNVLLSTPAPLYGGLALLTTGAFAFLFGVAFALASISSVPLCPLRLGSS	240
QY	241	ALTTOYGAAFWVTLATGYCLFLGAAVVSLOYVRPSALRTLIDQSAKDCSQERGSPIL	300
Db	241	ALTTOYGAAFWVTLATGYCLFLGAAVVSLOYVRPSALRTLIDQSAKDCSQERGSPIL	300
QY	301	GDPLHKQALPDLKCITTNL	320
Db	301	GDPLHKQALPDLKCITTNL	320
RESULT 5			
ADD19303	ID	ADD19303	standard; protein; 370 AA.
XX	AC	ADD19303;	
XX	DT	15-JAN-2004	(first entry)
DE	XX	Human secreted protein from gene 4 #2.	
KW	XX	human secreted protein; cytostatic; antibacterial; virucide;	
KW	XX	neuroprotective; gynaecological; gastrointestinal-Gen; cardiant;	
KW	XX	cardiovascular-Gen; nephrotropic; antiinflammatory; muscular-Gen;	
KW	XX	respiratory-Gen; immunosuppressive; cerebroprotective; vasotropic;	
KW	XX	nootropic; antiallergic; cancer; bacterial infection; viral infection;	
KW	XX	neural disorder; immune system disorder; blood disorder;	
KW	XX	muscular disorder; reproductive disorder; gastrointestinal disorder;	
KW	XX	pulmonary disorder; cardiovascular disorder; renal disorder;	
KW	XX	inflammatory disorder; proliferative disorder; human.	
OS	XX	Homo sapiens.	
XX	PN	WO2003052377-A2.	
XX	PD	26-JUN-2003.	
XX	PF	06-NOV-2002; 2002WO-US035606.	
XX	PR	07-NOV-2001; 2001US-0331046P.	
XX	PA	(HUMA-) HUMAN GENOME SCI INC.	
XX	PI	Rosen CA, Ruben SM;	
XX	DR	WPI; 2003-533050/50.	
XX	DR	N-PSDB; ADD19228.	
PT	XX	New isolated nucleic acids encoding signal transduction pathway component	
PT	XX	polypeptides, useful for diagnosing, treating, and/or preventing	
PT	XX	disorders, such as cancer, infections, cardiovascular and inflammatory	
PT	XX	diseases.	
XX	PS	Claim 11; SEQ ID NO 130; 554bp; English.	
XX	XX	The invention relates to an isolated nucleic acid molecule (cDNA)	
XX	XX	encoding a human secreted protein, representing one of 85 novel genes.	
XX	XX	Also included are recombinant vectors, host cells (expressing the	
XX	XX	protein), the secreted proteins (including their fragments, epitopes and	
XX	XX	homologues), an isolated antibody that binds specifically to the protein,	
XX	XX	diagnosing a pathological condition or susceptibility to a pathological	
XX	XX	condition (comprising determining the presence or absence of a mutation	
XX	XX	in the nucleic acid and diagnosing a condition based on the presence or	
XX	XX	absence of the mutation), diagnosing a pathological condition or	
XX	XX	susceptibility to a pathological condition (comprising determining the	
XX	XX	presence or amount of expression of the protein in a biological sample	
XX	XX	and diagnosing a condition based on the presence or amount of expression	

CC of the protein), preventing, treating or ameliorating a medical condition
CC by administering the nucleic acid or protein to a mammalian subject,
CC identifying a binding partner to the protein, the gene corresponding to
CC the cDNA sequence, and identifying an activity in a biological assay
CC (comprising expressing the nucleic acid in a cell, isolating the
CC supernatant, detecting an activity in a biological assay and identifying
CC the protein in the supernatant having the activity). The nucleic acids
CC and proteins display the following activities Cytostatic, antibacterial,
CC Virucide, Neuroprotective, Gynaecological, Gastrointestinal-Gen,
CC Cardiac, Cardiovascular-Gen, Nephrotropic, Antiinflammatory, Muscular-
CC Gen, Respiratory-Gen, Immunosuppressive, Cerebroprotective, Vasotropic,
CC Nootropic, Antiallergic. The methods and compositions of the present
CC invention are useful for diagnosing, treating, preventing and/or
CC prognosticating disorders related to the novel polypeptides, such as
CC cancer, bacterial or viral infections, and neural, immune system, blood,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal, inflammatory or proliferative disorders (many examples of these
CC diseases and disorders are given in the specification). The present
CC sequence represents a novel secreted protein of the invention.

Query Match	68.4%;	Score 219;	DB 7;	Length 370;
Best Local Similarity	99.7%;	Pred. No. 6.9e-203;		
Matches 319; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

QY	1	MTLWNGVLPEYPQPRHAGFSVPLLIVLVLPAALAAFLLLPGIRGSRMFWLVRLVLS	60
Db	51	MTLWNGVLPEYPQPRHAGFSVPLLIVLVLPAALAAFLLLPGIRGSRMFWLVRLVLS	110
QY	61	LFIGAIVAVHFSAEWFGVTNNTSYKAFSARVTARVGLIVLEGINITLTGTPTVHQL	120
Db	111	LFIGAIVAVHFSAEWFGVTNNTSYKAFSARVTARVGLIVGLXGINITLTGTPTVHQL	170
QY	121	NETIDYNEQFTWRLKENYAAEYANALEKGLPBPVLYLAEKTPSSPCGLYHQYHLAGHYA	180
Db	171	NETIDYNEQFTWRLKENYAAEYANALEKGLPBPVLYLAEKTPSSPCGLYHQYHLAGHYA	230
QY	181	SATLWVAFCFWLLSNVLLSTPAPLYGGLALLTGAFAFLFGVPAALASISSVPLCPLRLGSS	240
Db	231	SATLWVAFCFWLLSNVLLSTPAPLYGGLALLTGAFAFLFGVPAALASISSVPLCPLRLGSS	290
QY	241	ALTTOYGAAFWVTLATGVLCFLFGGAVVSLQYVRPSALRTLDDQSAKDCSQBERGSSPLIL	300
Db	291	ALTTOYGAAFWVTLATGVLCFLFGGAVVSLQYVRPSALRTLDDQSAKDCSQBERGSSPLIL	350
QY	301	GDPPLHKQAAALPDLKCITTNL 320	
Db	351	GDPPLHKQAAALPDLKCITTNL 370	

RESULT 6	
AA026326	
ID	AA026326 standard; protein; 262 AA.
XX	
AC	AA026326;
XX	
DT	12-SEP-2003 (revised)
DT	11-JAN-2001 (first entry)
XX	
DE	CASB618 protein.
XX	
KM	Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian,
KM	colon; autoimmune disease; HLA_A0201.
XX	
OS	Homo sapiens.
OS	Influenza virus.
OS	Chimeric.
XX	
PN	W0200053748-A2.
XX	
PD	14-SEP-2000.
XX	

PF 09-MAR-2000; 2000WO-EP002048.
 XX
 PR 11-MAR-1999; 99GB-00005607.
 PR 01-SEP-1999; 99GB-00020590.
 XX
 XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PI Bruck CEM, Cassart J, Coche T, Vinals Y De Bassolisc;
 XX
 DR WPI; 2000-572268/53.
 XX
 PT New human CASB618 polypeptide, useful as a vaccine for prophylactic and
 PT therapeutic treatment of cancers, particularly ovarian or colon cancer,
 PT autoimmune diseases and related conditions.
 XX
 PS Example 7; Page 63; 76pp; English.
 XX
 CC The present invention relates to epitopes of human CASB618 protein (see
 CC AAB26327 to AAB26399). The epitopes of AAB26327 to AAB26399 are useful in
 CC diagnosing the occurrence of tumour cells and in vaccines for prophylactic
 CC and therapeutic treatment of cancers, particularly ovarian or colon
 CC cancer, autoimmune diseases and related conditions. The present sequence
 CC is a protein derived from a chimeric gene constructed from the CASB618
 CC gene carrying deletions of the N-terminus and C-terminus, with the
 CC addition of an IFP (NSI DNA sequence encoding the N-terminal 1-81 amino
 CC acids of the NSI protein of Influenza virus) at the N-terminus, and a C-
 CC terminal histidine tail. The gene encoding the present protein was cloned
 CC and used to evaluate the effectiveness of tumour-specific antigens.
 CC (Updated on 12-SEP-2003 to standardise OS field)
 XX
 SQ Sequence 262 AA;
 Query Match 53.8%; Score 172; DB 3; Length 262;
 Best Local Similarity 100.0%; Pred. No. 1.4e-157;
 Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 75 EWFVGTVNTNTSYKAFSARVTVARVGLVGLGGINITLTGTPVHQINERTIDYNEQFTWRL 134
 DB 82 EWFVGTVNTNTSYKAFSARVTVARVGLVGLGGINITLTGTPVHQINERTIDYNEQFTWRL 141
 QY 135 KENYAAEYANALKEGLPDPVLYLAEKFTPSPCGLYHQYHLAGHYASATLWVAFCEFWLLS 194
 DB 142 KENYAAEYANALKEGLPDPVLYLAEKFTPSPCGLYHQYHLAGHYASATLWVAFCEFWLLS 201
 QY 195 NVLLSTPAPLYGGLALLTTGAFALFGVFPALASISSVPLCPRLGSSALTTQY 246
 DB 202 NVLLSTPAPLYGGLALLTTGAFALFGVFPALASISSVPLCPRLGSSALTTQY 253
 RESULT 7
 ID ADD19262 standard; protein; 103 AA.
 XX
 AC ADD19262;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Human secreted protein from gene 4.
 XX
 KW human secreted protein; cytostatic; antibacterial; virucide;
 KW neuroprotective; gynaecological; gastrointestinal-Gen; cardiant;
 KW cardiovascular-Gen; nephrotropic; antiinflammatory; muscular-Gen;
 KW respiratory-Gen; immunosuppressive; cerebroprotective; vasotropic;
 KW nootropic; antiallergic; cancer; bacterial infection; viral infection;
 KW neural disorder; immune system disorder; blood disorder;
 KW muscular disorder; reproductive disorder; gastrointestinal disorder;
 KW pulmonary disorder; cardiovascular disorder; renal disorder;
 KW inflammatory disorder; proliferative disorder; human.
 XX
 OS Homo sapiens.
 XX
 PN WO2003052377-A2.
 XX

PD 26-JUN-2003.
 XX
 XX 06-NOV-2002; 2002WO-US035606.
 XX
 PR 07-NOV-2001; 2001US-0331046P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2003-533050/50.
 DR N-PSDB; ADD19187.
 XX
 PT New isolated nucleic acids encoding signal transduction pathway component
 PT polypeptides, useful for diagnosing, treating, and/or preventing
 PT disorders, such as cancer, infections, cardiovascular and inflammatory
 PT diseases.
 XX
 PS Claim 11; SEQ ID NO 89; 554pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (cDNA)
 CC encoding a human secreted protein, representing one of 85 novel genes.
 CC Also included are recombinant vectors, host cells (expressing the
 CC protein), the secreted proteins (including their fragments, epitopes and
 CC homologues), an isolated antibody that binds specifically to the protein,
 CC diagnosing a pathological condition or susceptibility to a pathological
 CC condition (comprising determining the presence or absence of a mutation
 CC in the nucleic acid and diagnosing a condition based on the presence or
 CC absence of the mutation), diagnosing a pathological condition or
 CC susceptibility to a pathological condition (comprising determining the
 CC presence or amount of expression of the protein in a biological sample
 CC and diagnosing a condition based on the presence or amount of expression
 CC of the protein), preventing, treating or ameliorating a medical condition
 CC by administering the nucleic acid or protein to a mammalian subject,
 CC identifying a binding partner to the protein, the gene corresponding to
 CC the cDNA sequence, and identifying an activity in a biological assay
 CC (comprising expressing the nucleic acid in a cell, isolating the
 CC supernatant, detecting an activity in a biological assay and identifying
 CC the protein in the supernatant having the activity). The nucleic acids
 CC and proteins display the following activities: Cytostatic, antibacterial,
 CC Virucide, Neuroprotective, Gynaecological, Gastrointestinal-Gen,
 CC Cardiant, Cardiovascular-Gen, Nephrotropic, Antiinflammatory, Muscular-
 CC Gen, Respiratory-Gen, Immunosuppressive, Cerebroprotective, Vasotropic,
 CC Nootropic, Antiallergic. The methods and compositions of the present
 CC invention are useful for diagnosing, treating, preventing and/or
 CC prognosticating disorders related to the novel polypeptides, such as
 CC cancer, bacterial or viral infections, and neural, immune system, blood,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal, inflammatory or proliferative disorders (many examples of these
 CC diseases and disorders are given in the specification). The present
 CC sequence represents a novel secreted protein of the invention.
 XX
 SQ Sequence 103 AA;
 Query Match 15.3%; Score 49; DB 7; Length 103;
 Best Local Similarity 100.0%; Pred. No. 5.3e-39;
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTLNMGVLPFYPPRHAAGFSVPLLVLFVFLAASFLILPGRGHS 49
 DB 1 MTLNMGVLPFYPPRHAAGFSVPLLVLFVFLAASFLILPGRGHS 49
 RESULT 8
 ID AAG76780 standard; protein; 98 AA.
 XX
 AC AAG76780;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen protein SEQ ID NO:7544.
 XX

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.
XX
PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI; 2001-235357/24.
DR N-PSDB; AAH36185.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS Claim 11; Page 8959-8960; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patient's own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
SQ Sequence 98 AA;
QY
Query Match 5.9%; Score 19; DB 4; Length 98;
Best Local Similarity 100.0%; Pred. No. 5.2e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 50 RWFVLVRVLISLFIGAEIV 68
8 RWFVLVRVLISLFIGAEIV 26
RESULT 9
AAB66073
ID AAB66073 standard; protein; 112 AA.
XX
AC AAB66073;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human INTERCEPT 296 extracellular domain #2.
XX
KW TANGO protein; INTERCEPT protein; neurological disorder;
KW central nervous system; focal brain disorder; bipolar affective disorder;
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW neuropsychiatric; psychoactive substance use; anxiety.
XX

OS Homo sapiens.
XX
PN WO200077239-A2.
XX
PD 21-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US014858.
XX
PR 14-JUN-1999; 99US-00333159.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI McCarthy SA, Fraser CC, Sharp JD, Barnes TW;
XX
XX WPI; 2001-032313/04.
DR
XX
PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT screening assays and diagnostic assays and for the treatment of
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT disease.
XX
PS Claim 8; Page 331; 359pp; English.
XX
XX The present invention relates to TANGO or INTERCEPT proteins and coding
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC sequences are useful for the treatment of neurological disorders such as
CC central nervous system (CNS) disorders, CNS-related disorders, focal
CC brain disorders, global-diffuse cerebral disorders and other neurological
CC and cerebrovascular disorders. The CNS disorders include Alzheimer's
CC disease, senile dementia, Huntington's disease, amyotrophic lateral
CC sclerosis, Parkinson's, Gilles de la Tourette's syndrome, autonomic
CC function disorders such as hypertension and sleep disorders,
CC neuropsychiatric disorders, psychoactive substance use disorders,
CC anxiety, and bipolar affective disorder
XX
SQ Sequence 112 AA;
QY
Query Match 5.9%; Score 19; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 6e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 145 ALEKGLPDPVLYIAEKFTP 163
75 ALEKGLPDPVLYIAEKFTP 93
RESULT 10
ABO32727
ID ABO32727 standard; protein; 112 AA.
XX
AC ABO32727;
XX
DT 17-SEP-2003 (first entry)
XX
DE Secreted polypeptide-related protein #101.
XX
KW Human; TANGO; INTERCEPT; secreted polypeptide; immune disorder;
KW hormonal disorder; proliferative disorder; cancer; thyroid disorder;
KW diabetes; multiple sclerosis; lupus; neurological disorder; anaemia;
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KW myocardial infarction; congestive heart disease; blood platelet disorder;
KW thrombocytopenia; blood vessel; atherosclerosis; vasculitis.
XX
OS Homo sapiens.
XX
PN US2003022279-A1.
XX
PD 30-JAN-2003.
XX
PF 12-JAN-2001; 2001US-00759130.
XX
PR 14-JUN-1999; 99US-00333159.

PR 29-JUN-1999; 99US-00342364.
PR 10-SEP-1999; 99US-00393996.
PR 19-OCT-1999; 99US-00420707.
PR 07-JAN-2000; 2000US-00479249.
PR 27-APR-2000; 2000US-00559497.
PR 24-MAY-2000; 2000US-00578063.
PR 16-JUN-2000; 2000US-00596194.
PR 23-JUN-2000; 2000US-00602871.
PR 30-JUN-2000; 2000US-00608452.
XX
PA (FRASER/) FRASER C C.
PA (BARN/) BARNES T M.
PA (SHAR/) SHARP J D.
PA (KIRS/) KIRST S J.
PA (MYER/) MYERS P S.
PA (LEIB/) LEIBY K R.
PA (HOLT/) HOLTZMAN D A.
PA (MCCA/) MCCARTHY S A.
PA (WRIG/) WRIGHTON N.
PA (MACK/) MACKAY C R.
PA (GOOD/) GOODEARL A D J.
XX
PI Fraser CC, Barnes TM, Sharp JD, Kirst SJ, Myers PS, Leiby KR;
PI Holtzman DA, McCarthy SA, Wrighton N, Mackay CR, Goodearl ADJ;
XX
DR WPI; 2003-456290/43.
XX
PT New nucleic acid molecule encoding a secreted protein (e.g. TANGO 202,
PT TANGO 210 or INTERCEPT 217), useful for diagnosing, preventing or
PT treating disorders such as cancer, diabetes or atherosclerosis, and in
PT forensic biology.
XX
XX
PS Claim 9; SEQ ID NO 427; 482pp; English.
XX
XX The invention relates to secreted polypeptide-related proteins and
CC nucleic acids (TANGO and INTERCEPT proteins and nucleic acids). The
CC nucleic acids, proteins and antibodies specific to the proteins are
CC useful in screening assays, predictive medicine (e.g. diagnostic assays,
CC prognostic assays, monitoring clinical trials and pharmacogenetics) and
CC prophylactic and therapeutic methods. The sequences are used in
CC diagnosing, preventing or treating proliferative disorders (e.g.
CC cancers), hormonal disorders (e.g. diabetes or thyroid disorders), immune
CC disorders (e.g. multiple sclerosis or lupus), neurological disorders
CC (e.g. Alzheimer's disease or Parkinson's disease), cardiovascular
CC disorders (e.g. myocardial infarction or congestive heart disease), blood
CC platelet disorders (e.g. thrombocytopenia or anaemia) and disorders
CC involving blood vessels (e.g. atherosclerosis or vasculitis). The nucleic
CC acids may also be used in chromosome mapping, tissue typing and forensic
CC biology, and as surrogate markers. This sequence represents a secreted
CC polypeptide-related protein of the invention. Note: The sequence data for
CC this patent was obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 112 AA;

Query Match 5.9%; Score 19; DB 6; Length 112;
Best Local Similarity 100.0%; Pred. No. 6e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 ALEKGLPDPVLYLAEKFTP 163
Db 75 ALEKGLPDPVLYLAEKFTP 93

RESULT 11
ADA55266
ID ADA55266 standard; protein; 298 AA.
XX
AC ADA55266;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human protein, SEQ ID 2834.

XX
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease.
XX
OS Homo sapiens.
XX
EN EP1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-395539/38.
DR N-PSDB; ADA53627.
XX
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 14; SEQ ID NO 2834; 205pp; English.
XX
XX The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 298 AA;

Query Match 5.9%; Score 19; DB 6; Length 298;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 ALEKGLPDPVLYLAEKFTP 163
Db 100 ALEKGLPDPVLYLAEKFTP 118

RESULT 12
AAB66071
ID AAB66071 standard; protein; 343 AA.
XX
AC AAB66071;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human INTERCEPT 296.
XX
KW TANGO protein; INTERCEPT protein; neurological disorder;
KW central nervous system; focal brain disorder; bipolar affective disorder;
KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW neuropsychiatric; psychoactive substance use; anxiety.
XX
OS Homo sapiens.
XX
EN WO200077239-A2.
XX
PD 21-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US014858.

XX 14-JUN-1999; 99US-00333159.
PR (MILL-) MILLENNIUM PHARM INC.
PA Mccarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX WPI; 2001-032313/04.
DR N-PSDB; AAF45133, AAF45134.
XX TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT screening assays and diagnostic assays and for the treatment of
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT disease.
XX Claim 8; Fig 7; 359pp; English.
XX The present invention relates to TANGO or INTERCEPT proteins and coding
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC sequences are useful for the treatment of neurological disorders such as
CC central nervous system (CNS) disorders, CNS-related disorders, focal
CC brain disorders, global-diffuse cerebral disorders and other neurological
CC and cerebrovascular disorders. The CNS disorders include Alzheimer's
CC disease, senile dementia, Huntington's disease, amyotrophic lateral
CC sclerosis, Parkinson's, Gilles de la Tourette's syndrome, autonomic
CC function disorders such as hypertension and sleep disorders,
CC neuropsychiatric disorders, psychoactive substance use disorders,
CC anxiety, and bipolar affective disorder
XX Sequence 343 AA;
SQ
Query Match 5.9%; Score 19; DB 4; Length 343;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 145 ALEKGLPDPVLYIAEKFTP 163
DB 145 ALEKGLPDPVLYIAEKFTP 163
RESULT 13
ABU08370
ID ABU08370 standard; protein; 343 AA.
XX
AC ABU08370;
XX
DT 03-JUN-2003 (first entry)
XX
DE Amino acid sequence for human INTERCEPT 296.
XX
KM Human; INTERCEPT 296; TANGO; cellular process; tissue typing;
KM forensic biology; cellular protease activity; cell interaction;
KM development; blood disorder; haematopoietic cell-related disorder;
KM growth; cell proliferation; cell differentiation; gamma delta T-cell;
KM immune system; metabolic disorder; homeostatic disorder; anaemia;
KM developmental bone disorder; osteoporosis; bacterial infection;
KM thrombocytopaenia; renal failure; liver disease; cystic fibrosis;
KM obesity; cancer; bronchitis; asthma; emphysema; pulmonary oedema;
KM respiratory distress syndrome; osteopathic; antibacterial; anti-naemic;
KM thrombolytic; nephrotropic; antiobesity; hepatotropic; cytostatic;
KM antiinflammatory; antiasthmatic.
XX
OS Homo sapiens.
XX
PN US2002182675-A1.
XX
PD 05-DEC-2002.
XX
PF 25-OCT-2001; 2001US-00042431.
XX
PR 14-JUN-1999; 99US-00333159.
PR 24-MAY-2000; 2000US-00578063.

XX (MILL-) MILLENNIUM PHARM INC.
PA Mccarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX WPI; 2003-328617/31.
DR N-PSDB; ABX94107.
XX
PT New TANGO and INTERCEPT proteins, useful as modulating agents in
PT regulating a variety of cellular processes, in chromosome mapping, in
PT tissue typing, and in forensic biology.
XX Disclosure; Fig 7A-7C; 232pp; English.
XX The present invention relates to the isolation of novel TANGO or
CC INTERCEPT proteins, and the polynucleotide sequences encoding them. The
CC invention discloses sequences for human TANGO 202, TANGO 234, TANGO 265,
CC TANGO 273, TANGO 286, TANGO 294 and INTERCEPT 296. The sequences for
CC murine TANGO 202 and TANGO 273 are also provided. The TANGO polypeptides
CC and the polynucleotide sequences encoding them are useful as modulating
CC agents in regulating a variety of cellular processes. The polynucleotide
CC sequences are useful as primers or hybridisation probes for the detection
CC of nucleic acids encoding TANGO polypeptides, chromosome mapping, tissue
CC typing, and in forensic biology. TANGO 202 can be used to diagnose,
CC prevent or treat disorders relating to aberrant cellular protease
CC activity, inappropriate interaction of cells with mediators,
CC inappropriate development, and blood and haematopoietic cell-related
CC disorders. TANGO 234 can be used to modulate growth, proliferation,
CC survival, differentiation, and activity of gamma delta T-cells. TANGO 265
CC can be used to prevent, diagnose and treat disorders characterised by
CC aberrant organisation or development of a tissue or organ, and for
CC modulating differentiation of cells of the immune system. TANGO 273 is
CC useful for diagnosing, treating or preventing e.g. metabolic, homeostatic
CC and developmental bone disorders (e.g. osteoporosis), and bacterial
CC infection. TANGO 286 is useful for treating or preventing e.g. anaemia,
CC thrombocytopaenia, renal failure or liver disease. TANGO 294 is useful
CC for treating or preventing e.g. cystic fibrosis or obesity. INTERCEPT 296
CC is useful for diagnosing, treating or preventing e.g. cancers,
CC bronchitis, cystic fibrosis, asthma, emphysema, pulmonary oedema, or
CC adult and infant respiratory distress syndromes. The present sequence
CC represents human INTERCEPT 296
XX
SQ Sequence 343 AA;
Query Match 5.9%; Score 19; DB 6; Length 343;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 145 ALEKGLPDPVLYIAEKFTP 163
DB 145 ALEKGLPDPVLYIAEKFTP 163
RESULT 14
ABO32725
ID ABO32725 standard; protein; 343 AA.
XX
AC ABO32725;
XX
DT 17-SEP-2003 (first entry)
XX
DE Secreted polypeptide-related protein #100.
XX
KM Human; TANGO; INTERCEPT; secreted polypeptide; immune disorder;
KM hormonal disorder; proliferative disorder; cancer; thyroid disorder;
KM diabetes; multiple sclerosis; lupus; neurological disorder; anaemia;
KM Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KM myocardial infarction; congestive heart disease; blood platelet disorder;
KM thrombocytopenia; blood vessel; atherosclerosis; vasculitis.
XX
OS Homo sapiens.
XX
PN US2003022279-A1.

XX 30-JAN-2003.
PD
XX
PF 12-JAN-2001; 2001US-00759130.
XX
PR 14-JUN-1999; 99US-00333159.
PR 29-JUN-1999; 99US-00342364.
PR 10-SEP-1999; 99US-00393996.
PR 19-OCT-1999; 99US-00420707.
PR 07-JAN-2000; 2000US-00479249.
PR 27-APR-2000; 2000US-00559497.
PR 24-MAY-2000; 2000US-00578063.
PR 16-JUN-2000; 2000US-00596194.
PR 23-JUN-2000; 2000US-00602871.
PR 30-JUN-2000; 2000US-00608452.
XX
PA (FRAS/) FRASER C C.
PA (BARN/) BARNES T M.
PA (SHAR/) SHARP J D.
PA (KIRS/) KIRST S J.
PA (MYER/) MYERS P S.
PA (LEIB/) LEIBY K R.
PA (HOLT/) HOLTZMAN D A.
PA (MCCA/) MCCARTHY S A.
PA (WRIG/) WRIGHTON N.
PA (MACK/) MACKAY C R.
PA (GOOD/) GOODEARL A D J.
XX
PI Fraser CC, Barnes TM, Sharp JD, Kirst SJ, Myers PS, Leiby KR;
PI Holtzman DA, McCarthy SA, Wrighton N, Mackay CR, Goodearl ADJ;
XX
DR WPI; 2003-456290/43.
DR N-PSDB; ACD66786, ACD66787.
XX
PT New nucleic acid molecule encoding a secreted protein (e.g. TANGO 202,
PT TANGO 210 or INTERCEPT 217), useful for diagnosing, preventing or
PT treating disorders such as cancer, diabetes or atherosclerosis, and in
PT forensic biology.
XX
PS Claim 9; Fig 30A-30C; 482pp; English.
XX
CC The invention relates to secreted polypeptide-related proteins and
CC nucleic acids (TANGO and INTERCEPT proteins and nucleic acids). The
CC nucleic acids, proteins and antibodies specific to the proteins are
CC useful in screening assays, predictive medicine (e.g. diagnostic assays,
CC prognostic assays, monitoring clinical trials and pharmacogenetics) and
CC prophylactic and therapeutic methods. The sequences are used in
CC diagnosing, preventing or treating proliferative disorders (e.g.
CC cancers), hormonal disorders (e.g. diabetes or thyroid disorders), immune
CC disorders (e.g. multiple sclerosis or lupus), neurological disorders
CC (e.g. Alzheimer's disease or Parkinson's disease), cardiovascular
CC disorders (e.g. myocardial infarction or congestive heart disease), blood
CC platelet disorders (e.g. thrombocytopenia or anaemia) and disorders
CC involving blood vessels (e.g. atherosclerosis or vasculitis). The nucleic
CC acids may also be used in chromosome mapping, tissue typing and forensic
CC biology, and as surrogate markers. This sequence represents a secreted
CC polypeptide-related protein of the invention. Note: The sequence data for
CC this patent was obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 343 AA;

Query Match 5.9%; Score 19; DB 6; Length 343;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 ALEKGLPDPVLYLAEKFTP 163
Db 145 ALEKGLPDPVLYLAEKFTP 163

RESULT 15
ABG03520

ID ABG03520 standard; protein; 445 AA.
XX
AC ABG03520;
XX
DT 13-FEB-2002 (First entry)
XX
DE Novel human diagnostic protein #3511.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS67707.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 33879; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 445 AA;

Query Match 5.9%; Score 19; DB 4; Length 445;
Best Local Similarity 100.0%; Pred. No. 2.3e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 ALEKGLPDPVLYLAEKFTP 163
Db 168 ALEKGLPDPVLYLAEKFTP 186

Search completed: February 23, 2004, 19:33:18
Job time : 60 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 21:58:00 ; Search time 5685.97 Seconds

(without alignments)
10984.465 Million cell updates/sec

Title: US-09-936-456-1

Perfect score: 1441

Sequence: 1 aaagtaacgctacagacag.....aaaaaaaaaaaaaa 1441

Scoring table: OLIGO_NUC

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1441	100.0	1441	6	AX035346	AX035346 Sequence
2	976	67.7	1740	9	HSM806241	BX537581 Homo sapi
3	425	29.5	498	6	AX035348	AX035348 Sequence
4	405	28.1	166937	2	AC012255	AC012255 Homo sapi
5	405	28.1	181312	9	AC091117	AC091117 Homo sapi
6	392	27.2	2684	6	AX714511	AX714511 Sequence
7	392	27.2	2684	9	AK056896	AK056896 Homo sapi
8	392	27.2	156534	2	AC009700	AC009700 Homo sapi
9	392	27.2	171444	2	AC087790	AC087790 Homo sapi
10	354	24.6	156534	2	AC009700	AC009700 Homo sapi
11	217	15.1	217	9	HS183H12F	Z57552 H. sapiens C
12	166	11.5	1923	9	HS183A12R	BC029819 Homo sapi
13	136	9.4	218	9	BC004161	Z59954 H. sapiens C
14	39	2.7	1725	9	BC003157	BC004161 Homo sapi
15	38	2.6	1260	9	BC017580	BC003157 Homo sapi
16	38	2.6	2165	6	BD275921	BC017580 Homo sapi
17	38	2.6	2384	6	AX780522	BD275921 Uncouplin
18	38	2.6	2384	6	AX780522	AX780522 Sequence
19	38	2.6	3393	9	BC036370	AX780523 Sequence
20	38	2.6	5642	6	AX345771	BC036370 Homo sapi
21	38	2.6	10480	6	AX347103	AX345771 Sequence
22	38	2.6	174790	2	AC134964	AX347103 Sequence
23	38	2.6	180059	2	AC023928	AC134964 Carls fam
24	38	2.6	180763	2	AC067826	AC023928 Homo sapi
25	38	2.6	185551	2	AC024386	AC067826 Homo sapi
26	38	2.6	185551	2	AC024386	AC024386 Homo sapi
27	38	2.6	1426	10	AF175224	AC015683 Homo sapi
28	37	2.6	1640	9	BC050738	AF175224 Rattus no
29	37	2.6	1678	9	HSM802263	BC050738 Homo sapi
30	37	2.6	1724	9	BC003101	AL137530 Homo sapi
31	37	2.6	2645	3	AY118446	BC003101 Homo sapi
32	37	2.6	3734	9	HSM807535	AY118446 Drosophil
33	37	2.6	3989	9	HSM805415	BX647390 Homo sapi
34	37	2.6	3990	9	HSM805416	AL834162 Homo sapi
35	37	2.6	135628	2	AC048357	AL834162 Homo sapi
36	37	2.6	153166	2	AC048357	AC067764 Homo sapi
37	37	2.6	190371	2	AC116830	AC048357 Homo sapi
38	37	2.6	349980	6	AX344557	AC116830 Mus muscu
39	36	2.5	1059	3	AF036416	AX344557 Sequence
40	36	2.5	1407	10	BC013450	AF036416 Trypanoso
41	36	2.5	1603	10	BC024516	BC013450 Mus muscu
42	36	2.5	1713	9	BC003624	BC024516 Mus muscu
43	36	2.5	2155	9	BC015749	BC003624 Homo sapi
44	36	2.5	2255	6	AR168373	BC015749 Homo sapi
45	36	2.5	2255	6	AR168373	AR168373 Sequence

ALIGNMENTS

RESULT 1
AX035346 1441 bp DNA linear PAT 15-NOV-2000
LOCUS AX035346 Sequence 1 from Patent WO0053748.
DEFINITION AX035346
ACCESSION AX035346
VERSION AX035346.1 GI:11191064
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Bruck,C.E., Vinals,Y.D., Coche,T. and Cassart,J.P.
TITLE Novel compounds
JOURNAL Patent: WO 0053748-A 1 14-SEP-2000;

BRUCK CLAUDE ELVIRE MARIE (BE) ; SMITHLINE BECHAM BIOLOG (BE) ;
 VINALS Y DE BASSOLS CARLOTA (BE) ; COCHE THIERRY (BE) ; CASSART
 JEAN POL (BE)

FEATURES
 source Location/Qualifiers

1..1441
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 1441; DB 6; Length 1441;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTACGGCTACAGACAGTGAAGAAATAGTTGCTCGCCGCTAGAAAACTGTGCG 60
 DB 1 AAAGTACGGCTACAGACAGTGAAGAAATAGTTGCTCGCCGCTAGAAAACTGTGCG 60

QY 61 GTACCAACCCAGAGCGCTTGAGAGAGAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 120
 DB 61 GTACCAACCCAGAGCGCTTGAGAGAGAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 120

QY 121 GGACTCAGACTTCACAGAGCCCACTCGCTCCAGCTTGTACGCAAGAGAGAGAGAGAG 180
 DB 121 GGACTCAGACTTCACAGAGCCCACTCGCTCCAGCTTGTACGCAAGAGAGAGAGAGAG 180

QY 181 GCGCTCTCCCGCGCTCCAGAGAGAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTG 240
 DB 181 GCGCTCTCCCGCGCTCCAGAGAGAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTG 240

QY 241 ACTCGCGCGCGCTGACAGATGACCTGTGGAACGCGCTACTGCTTTTACCCCGAGCCC 300
 DB 241 ACTCGCGCGCGCTGACAGATGACCTGTGGAACGCGCTACTGCTTTTACCCCGAGCCC 300

QY 301 CGGCATGCGCGAGCGCTTCAAGCGCTTCACTGCTCATGCTTATTTAGTTTGGCTCTA 360
 DB 301 CGGCATGCGCGAGCGCTTCAAGCGCTTCACTGCTCATGCTTATTTAGTTTGGCTCTA 360

QY 361 GCAGCAAGCTTCTGCTCATCTTGGCGGGGATCCGTGGCCACTCGCGCTGTTTGGTTG 420
 DB 361 GCAGCAAGCTTCTGCTCATCTTGGCGGGGATCCGTGGCCACTCGCGCTGTTTGGTTG 420

QY 421 GTGAGAGTTCTTCTCAGTCTGTTTCATAGGGCGAGAAATTGCTGTGACATTCAGTGCA 480
 DB 421 GTGAGAGTTCTTCTCAGTCTGTTTCATAGGGCGAGAAATTGCTGTGACATTCAGTGCA 480

QY 481 GAATGTTCTGTTGTTAGTGAACACCAACACATCTTCAAAAGCCTTCAAGCGAGCGCGC 540
 DB 481 GAATGTTCTGTTGTTAGTGAACACCAACACATCTTCAAAAGCCTTCAAGCGAGCGCGC 540

QY 541 GTTACAGCCCCGTGTCGCTGCTCGTGGGCTTGAGGGGCTTAAATTACACTCACAGGG 600
 DB 541 GTTACAGCCCCGTGTCGCTGCTCGTGGGCTTGAGGGGCTTAAATTACACTCACAGGG 600

QY 601 ACCCGAGTGCATCAGCTGAACGAGACCATTTGACTACACGAGAGTTCACTGAGCGTCTG 660
 DB 601 ACCCGAGTGCATCAGCTGAACGAGACCATTTGACTACACGAGAGTTCACTGAGCGTCTG 660

QY 661 AAAGAGATTACGCGCGGAGTAGACGCAACGACCTGGAGAGAGGGGCTCCCGAGCCAGTG 720
 DB 661 AAAGAGATTACGCGCGGAGTAGACGCAACGACCTGGAGAGAGGGGCTCCCGAGCCAGTG 720

QY 721 CTCTACCTGGCGGAGAGTTACACCGAGTAGCCCTTGGCGCTGTACACAGAGTACAC 780
 DB 721 CTCTACCTGGCGGAGAGTTACACCGAGTAGCCCTTGGCGCTGTACACAGAGTACAC 780

QY 781 CTGGCGGAGACTAGCGCTCGGCGAGCTATGGTGGCGTTCTGCTTCTGCTCCTCTCC 840
 DB 781 CTGGCGGAGACTAGCGCTCGGCGAGCTATGGTGGCGTTCTGCTTCTGCTCCTCTCC 840

QY 841 AACGTGCTGCTCTCCAGCGCGCGCGCTCTACGAGAGGCTGCACTGTGACCAACCGGA 900
 DB 841 AACGTGCTGCTCTCCAGCGCGCGCGCTCTACGAGAGGCTGCACTGTGACCAACCGGA 900

QY 901 GCCTTCGCGCTCTTCGGGGTCTTCGCTTGGCCCTCCATCTCTAGCGTGCCTGCCCC 960
 DB 901 GCCTTCGCGCTCTTCGGGGTCTTCGCTTGGCCCTCCATCTCTAGCGTGCCTGCCCC 960

QY 961 CTCGCGCTAGGCTCTCCGCGCTACACACTCAGTACGGCGCGCTTCTGGTACGCTG 1020
 DB 961 CTCGCGCTAGGCTCTCCGCGCTACACACTCAGTACGGCGCGCTTCTGGTACGCTG 1020

QY 1021 GCAACCGCGCTCTGCTCTCTCTCGAGAGGGCGGTGTGAGTCTCCAGTATGTTCCG 1080
 DB 1021 GCAACCGCGCTCTGCTCTCTCTCGAGAGGGCGGTGTGAGTCTCCAGTATGTTCCG 1080

QY 1081 CCCAGCGCTCTTCCAGCCCTTCTGACCAAGCGCCAAAGAGACTGACCCAGAGAGAGG 1140
 DB 1081 CCCAGCGCTCTTCCAGCCCTTCTGACCAAGCGCCAAAGAGACTGACCCAGAGAGAGG 1140

QY 1141 GGCTCAGCTCTTATCTCGGCGAGCCCACTGACAGAGAGCGCTCTCCAGACTTAAAA 1200
 DB 1141 GGCTCAGCTCTTATCTCGGCGAGCCCACTGACAGAGAGCGCTCTCCAGACTTAAAA 1200

QY 1201 TGTATCACCATAACCTGTGAGGGGGACCCCAATCTGGAATCTCTCCCGCTGGGACAT 1260
 DB 1201 TGTATCACCATAACCTGTGAGGGGGACCCCAATCTGGAATCTCTCCCGCTGGGACAT 1260

QY 1261 CGCAGCGCGGAGAGCAGTGCCTCCGCGCAGGCTGCGGCGAGAGAGCTCCAGAGGGCACTG 1320
 DB 1261 CGCAGCGCGGAGAGCAGTGCCTCCGCGCAGGCTGCGGCGAGAGAGCTCCAGAGGGCACTG 1320

QY 1321 AGCGCTGTGCGCGAGGCGCTCGGACATCCGAGAGGCAAGGAAAGTCTCTGGGGCGA 1380
 DB 1321 AGCGCTGTGCGCGAGGCGCTCGGACATCCGAGAGGCAAGGAAAGTCTCTGGGGCGA 1380

QY 1381 TCTGTAATATAACCTTTTCTTTTCTTTTAAATAAAAAAAAAAAAAAAAAAAAAA 1440
 DB 1381 TCTGTAATATAACCTTTTCTTTTCTTTTAAATAAAAAAAAAAAAAAAAAAAAAA 1440

QY 1441 A 1441
 DB 1441 A 1441

RESULT 2
 HSM806241 1740 bp mRNA linear PRI 17-JUN-2003
 LOCUS HSM806241
 DEFINITION Homo sapiens mRNA; cDNA DKFZp686C04213 (from clone DKFZp686C04213).
 ACCESSION BX537581
 VERSION BX537581.1 GI:31873659
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)

REFERENCE
 AUTHORS Bloeker, H., Boecker, M., Mewes, H.W., Weil, B., Amid, C., Oesanger, A.,
 Fobo, G., Han, M. and Wiemann, S.
 TITLE Direct Submission
 JOURNAL Submitted (17-JUN-2003) MIPS, Ingolstaedter Landstr.1, D-85764
 Neuberberg, GERMANY

COMMENT
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by GBF (National Research Centre for Biotechnology Ltd.,
 Braunschweig/Germany) within the CDNA sequencing consortium of the
 German Genome Project.

This clone (DKFZp686C04213) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further
 information about the clone and the sequencing project is available
 at <http://mips.gsf.de/proj/cdna/>.

FEATURES
 source Location/Qualifiers
 1..1740
 /organism="Homo sapiens"
 /mol_type="mRNA"

/db_xref="taxon:9606"
/map="15q15"
/clone="DKFZp686C04213"
/issue_type="human rectum tumor"
/clone_lib="686 (synonym: h1cc3). Vector pSPORT1_Sfi; host
DH10B; sites SfiI + SfiIB"
/dev_stage="adult"
1694.1699
polyA_signal
polyA_site
1720

Query Match 67.7%; Score 976; DB 9; Length 1740;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1026; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 405 GCGCTGTTTGGTTGGTAGAGTCTCTCAGTCTGTTCAATAGCGCAGAAATTGTGGC 464
DB 712 GCGCTGTTTGGTTGGTAGAGTCTCTCAGTCTGTTCAATAGCGCAGAAATTGTGGC 771
QY 465 TGTGACTTTCAGTGCAGATGTTCTGGGTACAGTGAACACCAACACATCTCAAAAGC 524
DB 772 TGTGACTTTCAGTGCAGATGTTCTGGGTACAGTGAACACCAACACATCTCAAAAGC 831
QY 525 CTTACGCGCAGCGCGCTTACAGCCGCTGTGGTCTGCTGCTGGGCTTGAGGGCATTAA 584
DB 832 CTTACGCGCAGCGCGCTTACAGCCGCTGTGGTCTGCTGCTGGGCTTGAGGGCATTAA 891
QY 585 TATTACACTCAGAGGAGCCCAAGTGCATCAGCTGAACGAGACCATTTACTACACGAGCA 644
DB 892 TATTACACTCAGAGGAGCCCAAGTGCATCAGCTGAACGAGACCATTTACTACACGAGCA 951
QY 645 GTTACCTGCGCTCTGAAGAAGATTACGCCCGGAGTACGCCAAGCACTGAGAGAGG 704
DB 952 GTTACCTGCGCTCTGAAGAAGATTACGCCCGGAGTACGCCAAGCACTGAGAGAGG 1011
QY 705 GCTGCCGAGCCAGTGTCTTACCTGGCGAGAAAGTTCAACCGAGTAGCCCTTGGCGCT 764
DB 1012 GCTGCCGAGCCAGTGTCTTACCTGGCGAGAAAGTTCAACCGAGTAGCCCTTGGCGCT 1071
QY 765 GTACCAACAGTACCACTGGCGGAGCACTACGCTCGGCCACGCTATGGGTGGCTTCTG 824
DB 1072 GTACCAACAGTACCACTGGCGGAGCACTACGCTCGGCCACGCTATGGGTGGCTTCTG 1131
QY 825 CTTCTGCTCTCTCCAGTGTCTCTCCACGCGCGCGCTCTCTACGAGGCGCTGGC 884
DB 1132 CTTCTGCTCTCTCCAGTGTCTCTCCACGCGCGCGCTCTCTACGAGGCGCTGGC 1191
QY 885 ACTGCTGACACCGAGCGCTTGGCGCTTGGGGCTTGGCTTGGCTCTCTAG 944
DB 1192 ACTGCTGACACCGAGCGCTTGGCGCTTGGGGCTTGGCTTGGCTCTCTAG 1251
QY 945 CGTGCGCTGCGCGCTCGCTAGGCTCTCGCGCTCACCACCTAGTACGGCGCGC 1004
DB 1252 CGTGCGCTGCGCGCTCGCTAGGCTCTCGCGCTCACCACCTAGTACGGCGCGC 1311
QY 1005 CTTCTGAGTACGCTGCAACCGCGCTCTCTGCTCTCTCGAGGGCGCGCTGGTGG 1064
DB 1312 CTTCTGAGTACGCTGCAACCGCGCTCTCTGCTCTCTCGAGGGCGCGCTGGTGG 1371
QY 1065 TCTCCAGTATGTTGGCCAGCGCTCTTGGCACCCCTTGTGACCAAGCGCCAAAGACTG 1124
DB 1372 TCTCCAGTATGTTGGCCAGCGCTCTTGGCACCCCTTGTGACCAAGCGCCAAAGACTG 1431
QY 1125 CAGCCAGGAGAGGGGGCTCACTCTTATCTCTGGGACCCACTGCACAAGAGGCGCGC 1184
DB 1432 CAGCCAGGAGAGGGGGCTCACTCTTATCTCTGGGACCCACTGCACAAGAGGCGCGC 1491
QY 1185 TCTCCAGACTTAAATGTATCAACCACTAAGCTGTAGAGGGGAGCCCAATCTGGACTCTT 1244
DB 1492 TCTCCAGACTTAAATGTATCAACCACTAAGCTGTAGAGGGGAGCCCAATCTGGACTCTT 1551
QY 1245 CCCCCCTTGGGACATCGCAGGCGCGGAGACAGTCCCGCCAGGCTTGGGCGAGAGAGC 1304

DB 1552 CCCCCCTTGGGACATCGCAGGCGCGGAGACAGTCCCGCCAGGCTTGGGCGAGAGAGC 1611
QY 1305 TCCAGGAGGCACTGAGCGCTGTGGCGGAGGCGCTCGGACATCGCAGGACCAAGGGA 1364
DB 1612 TCCAGGAGGCACTGAGCGCTGTGGCGGAGGCGCTCGGACATCGCAGGACCAAGGGA 1671
QY 1365 AAGTCTCTGGGGCGATCTGTAAATAAACCCTTTTCTTTTGTGTTTTTAAAAAAA 1424
DB 1672 AAGTCTCTGGGGCGATCTGTAAATAAACCCTTTTCTTTTGTGTTTTTAAAAAAA 1731
QY 1425 AAAAAA 1431
DB 1732 AAAAAA 1738

RESULT 3
AX035348
LOCUS AX035348 498 bp DNA linear PAT 15-NOV-2000
DEFINITION Sequence 3 from Patent WO0053748.
ACCESSION AX035348
VERSION AX035348.1 GI:11191065
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Bruck,C.E., Vinals,Y.D., Coche,T. and Cassart,J.P.
TITLE Novel compounds
JOURNAL Patent: WO 0053748-A 3 14-SEP-2000;
BRUCK CLAUDE ELVIRE MARIE (BE) ; SMITHKLINE BEECHAM BIOLOG (BE) ;
VINALS Y DE BASSOLS CARLOTTA (BE) ; COCHE THIERRY (BE) ; CASSART
JEAN POL (BE)
FEATURES
source location/Qualifiers
1..498
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 29.5%; Score 425; DB 6; Length 498;
Best Local Similarity 100.0%; Pred. No. 5.8e-224; Indels 0; Gaps 0;
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 998 GCGCGCGCTTCTGGGTACGCTGGCAACCGCGCTCTGTGCTCTTCTCGAGGGGCGC 1057
DB 62 GCGCGCGCTTCTGGGTACGCTGGCAACCGCGCTCTGTGCTCTTCTCGAGGGGCGC 121
QY 1058 TGGTAGTCTCCAGTATGTTGGGCCCAAGCGCTTCTCGACCTTCTGACCAAAAGCGCA 1117
DB 122 TGGTAGTCTCCAGTATGTTGGGCCCAAGCGCTTCTCGACCTTCTGACCAAAAGCGCA 181
QY 1118 AGGACTGCAAGGAGAGAGGGGCTCACCTCTTATCTCTGGCGAACCACCTGACCAAGC 1177
DB 182 AGGACTGCAAGGAGAGAGGGGCTCACCTCTTATCTCTGGCGAACCACCTGACCAAGC 241
QY 1178 AGGCGCTCTCCAGACTTAAATGTATCACTAAGCTGTGAGGGGAGCCCAATCTGG 1237
DB 242 AGGCGCTCTCCAGACTTAAATGTATCACTAAGCTGTGAGGGGAGCCCAATCTGG 301
QY 1238 ACTCTTCCCGCGCTTGGGACATCGCAGCGCGGGAAGAGAGTCCCGCGCAGGCTGGGCA 1297
DB 302 ACTCTTCCCGCGCTTGGGACATCGCAGCGCGGGAAGAGAGTCCCGCGCAGGCTGGGCA 361
QY 1298 GAGAGCTCCAGAGAGGCACTGAGCGCTGGCGCGAGGCGCTCGGACATCCGAGGCA 1357
DB 362 GAGAGCTCCAGAGAGGCACTGAGCGCTGGCGCGAGGCGCTCGGACATCCGAGGCA 421
QY 1358 CCAGGAAAAGTCTCTGGGCGATCTGTAAATAAACCCTTTTCTTTTGTGTTTTTAAAA 1417
DB 422 CCAGGAAAAGTCTCTGGGCGATCTGTAAATAAACCCTTTTCTTTTGTGTTTTTAAAA 481
QY 1418 AAAAA 1422

```

|||||
Db      482 AAAA 486

RESULT 4
AC012255
LOCUS   AC012255
DEFINITION Homo sapiens chromosome 15 clone RP11-109D20 map 15, WORKING DRAFT
SEQUENCE AC012255
AC012255.4 GI:7684440
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE   Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 166937)
TITLE    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL  Homo sapiens chromosome 15, clone RP11-109D20
REFERENCE
AUTHORS 2 (bases 1 to 166937)
TITLE    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Margulis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tefaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 3, 2000 this sequence version replaced gi:7209928.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L3737
Center clone name: 109.D.20

----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 153800 bases at least Q40
Consensus quality: 160606 bases at least Q30
Consensus quality: 163275 bases at least Q20
Insert size: 164837; sum-of-contigs
Quality coverage: 4.0 in Q20 bases; sum-of-contigs

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 2087: contig of 2087 bp in length
* 2088 2187: gap of 100 bp
* 2188 3626: contig of 1439 bp in length
* 3627 3726: gap of 100 bp
* 3727 4845: contig of 1119 bp in length

```

```

FEATURES
source
* 4846 4945: gap of 100 bp
* 4946 6702: contig of 1757 bp in length
* 6703 6802: gap of 100 bp
* 6803 8326: contig of 1524 bp in length
* 8327 8426: gap of 100 bp
* 8427 10494: contig of 2068 bp in length
* 10495 10594: gap of 100 bp
* 10595 13215: contig of 2621 bp in length
* 13216 13315: gap of 100 bp
* 13316 16451: contig of 3136 bp in length
* 16452 16551: gap of 100 bp
* 16552 20843: contig of 4292 bp in length
* 20844 20943: gap of 100 bp
* 20944 24940: contig of 3997 bp in length
* 24941 25041: gap of 100 bp
* 25041 29901: contig of 4861 bp in length
* 29902 30001: gap of 100 bp
* 30002 34189: contig of 4188 bp in length
* 34190 34289: gap of 100 bp
* 34290 38386: contig of 4097 bp in length
* 38387 38486: gap of 100 bp
* 38487 45827: contig of 7341 bp in length
* 45828 45927: gap of 100 bp
* 45928 51729: contig of 5802 bp in length
* 51730 51829: gap of 100 bp
* 51830 60123: contig of 8294 bp in length
* 60124 60223: gap of 100 bp
* 60224 73532: contig of 13309 bp in length
* 73533 73632: gap of 100 bp
* 73633 87433: contig of 13801 bp in length
* 87434 87533: gap of 100 bp
* 87534 101708: contig of 14175 bp in length
* 101709 101808: gap of 100 bp
* 101809 121193: contig of 19385 bp in length
* 121194 121293: gap of 100 bp
* 121294 139051: contig of 17758 bp in length
* 139052 139151: gap of 100 bp
* 139152 166937: contig of 27786 bp in length.

1. 166937
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-109D20"
/clone_lib="RPCI-11 Human Male BAC"
1. 2087
/note="assembly_fragment"
2188. 3626
/note="assembly_fragment"
3727. 4845
/note="assembly_fragment"
4946. 6702
/note="assembly_fragment"
6803. 8326
/note="assembly_fragment"
8427. 10494
/note="assembly_fragment"
10595. 13215
/note="assembly_fragment"
13316. 16451
/note="assembly_fragment"
16552. 20843
/note="assembly_fragment"
20944. 24940
/note="assembly_fragment"
25041. 29901
/note="assembly_fragment"
30002. 34189
/note="assembly_fragment"
34290. 38386
/note="assembly_fragment"

```



```
misc_feature /note="sequence data generated from subcloned PCR product"
53030. .53480
/note="sequence data generated from subcloned PCR product"
unsure 86982. .86983
/note="low quality data"
misc_feature 95610. .96030
/note="sequence data generated from subcloned PCR product"
misc_feature 96830. .97420
/note="sequence data generated from subcloned PCR product"
misc_feature 98260. .98420
/note="single clone coverage"
unsure 98850. .98852
/note="low quality data"
unsure 100981. .101003
/note="low quality data"
misc_feature 174237. .179372
/note="overlap with RP11-276K9, AC087790. This sequence
data was added to establish the overlap between the BACs
in the tiling path"
179373. .181312
/note="overlap with CTD-2651B20, AC051619. This data is
added to give overlap for long range contiguity."
```

ORIGIN

Query Match	28.1%;	Score 405;	DB 9;	Length 181312;
Best Local Similarity	100.0%;	Pred. No. 7.4e-213;		
Matches 405;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	AAAGTAA	CGGCTACAGACAGTGA	GAAATAGTTTCGCTCGCCGCGCTAGAAAAA	CTGTGC	60
Db	165016	AAAGTAACGGCTACAGACAGTGA	GAAATAGTTTCGCTCGCCGCGCTAGAAAAA	CTGTGC	165075	
QY	61	GTA	CCAA	CCCCCAGAGCGTTGAGAGCAGCCCCACCTCCACGCTTCTTAACGGAGAGGTGCA	120	
Db	165076	GTA	CCAA	CCCCCAGAGCGTTGAGAGCAGCCCCACCTCCACGCTTCTTAACGGAGAGGTGCA	165135	
QY	121	GGA	CTCAGACTTCA	CCAGCCCACTCCGCTCCCAAGCCTTGTA	CGCAAGAGACGCCAAGAC	180
Db	165136	GGA	CTCAGACTTCA	CCAGCCCACTCCGCTCCCAAGCCTTGTA	CGCAAGAGACGCCAAGAC	165195
QY	181	GCG	CTCTCCCGCGCTCCAGGACAGCCCCAGCTTGCTGCTTGCTGCGCCGCTGCGTGACG	240		
Db	165196	GCG	CTCTCCCGCGCTCCAGGACAGCCCCAGCTTGCTGCTTGCTGCGCCGCTGCGTGACG	165255		
QY	241	ACT	CGGCGCGCGCTGACGATGACCTGTGGAACGCGCTACTGCTTTTAA	CCCCCAGCCC	300	
Db	165256	ACT	CGGCGCGCGCTGACGATGACCTGTGGAACGCGCTACTGCTTTTAA	CCCCCAGCCC	165315	
QY	301	CGG	CATGCCCGCAGAGCTTCAGCGTTCCACTGCTCATCGTTATTCTAGTGTTTTGGCTCTA	360		
Db	165316	CGG	CATGCCCGCAGAGCTTCAGCGTTCCACTGCTCATCGTTATTCTAGTGTTTTGGCTCTA	165375		
QY	361	GCA	GCAAGCTTCTGCTCATCTTTGCGGGGATCCGTGGCCACTCG	405		
Db	165376	GCA	GCAAGCTTCTGCTCATCTTTGCGGGGATCCGTGGCCACTCG	165420		

RESULT 6	AX714511/c	2684 bp	DNA	linear	PAT 15-APR-2003
LOCUS	AX714511				
DEFINITION	Sequence 1195 from Patent EP1293569.				
ACCESSION	AX714511				
VERSION	AX714511.1	GI:29889464			
KEYWORDS					
SOURCE					
ORGANISM	Homo sapiens (human)				
	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1				
AUTHORS	Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masuko, Y.				

```

TITLE      Full-length cDNAs
JOURNAL    Patent: EP 1293569-A 1195 19-MAR-2003;
            Helix Research Institute (JP) ; Research Association for
            Biotechnology (JP)
FEATURES   location/Qualifiers
            1..2684
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

```

ORIGIN

Query Match	27.2%;	Score 392;	DB 6;	Length 2684;
Best Local Similarity	100.0%;	Pred. No. 1.2e-205;		
Matches 392;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1027	GGCGTCTGTGCTCTCTTCTCTCGAGGGGCGGTGTTAGTCTCCAGTATGTTGGCCCA	1086
Db	2343	GGCGTCTGTGCTCTCTTCTCTCGAGGGGCGGTGTTAGTCTCCAGTATGTTGGCCCA	2284
QY	1087	GCTCTTCGCAACCTTCTTGGAACCAAGCGCCAAAGACTGACAGCCAGAGAGGGGCTCA	1146
Db	2283	GCTCTTCGCAACCTTCTTGGAACCAAGCGCCAAAGACTGACAGCCAGAGAGGGGCTCA	2224
QY	1147	CCTCTTATCTCGGCGCAACCACTGCACAGCAGGCGCTCTCCAGACTTAAATGTATC	1206
Db	2223	CCTCTTATCTCGGCGCAACCACTGCACAGCAGGCGCTCTCCAGACTTAAATGTATC	2164
QY	1207	ACCACTAACCTGTGAGGGGGACCCAATCTGACTCTTCCCGCCTTGGACATCGCAGG	1266
Db	2163	ACCACTAACCTGTGAGGGGGACCCAATCTGACTCTTCCCGCCTTGGACATCGCAGG	2104
QY	1267	CCGGGAAGCAGTCCCGCCAGGCTTGCGCCAGAGAGCTCCAGGAAGGCCACTGAGCGCT	1326
Db	2103	CCGGGAAGCAGTCCCGCCAGGCTTGCGCCAGAGAGCTCCAGGAAGGCCACTGAGCGCT	2044
QY	1327	GCTGCGCGAGGCTCTCGACATCCGACAGCACCAAGGAAAGTCTCCTGGGCGATCTGTA	1386
Db	2043	GCTGCGCGAGGCTCTCGACATCCGACAGCACCAAGGAAAGTCTCCTGGGCGATCTGTA	1984
QY	1387	AATAAACCTTTTTTCTTTTGTTTTTTAAAAA	1418
Db	1983	AATAAACCTTTTTTCTTTTGTTTTTTAAAAA	1952

RESULT 7	AK056896/c	2684 bp	mRNA	linear	PRI 01-AUG-2002
LOCUS	AK056896				
DEFINITION	Homo sapiens cDNA FLJ32334 fis, clone PROST2005426.				
ACCESSION	AK056896				
VERSION	AK056896.1				
KEYWORDS	oligo capping; fis (full insert sequence).				
SOURCE	Homo sapiens (human)				

REFERENCE
1
Ishibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S.,

Ono, Y., Horuta, T., Hiraoka, S., Murakawa, K., Takiguchi, S., Kusanagi, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M., Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, T., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuko, Y., Nagai, K. and Isogai, T.

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 2684)
Isogai, T., Otsuki, T. and Sugiyama, T.
Direct Submission
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

COMMENT

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES

source

Location/Qualifiers

1..2684

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="PROST2005426"

/tissue type="prostate"

/clone_id="PROST2"

/note="cloning vector: pME18SFL3"

56..952

/note="unnamed protein product"

/codon_start=1

/protein_id="BAB71304.1"

/db_xref="GI:16552420"

/translation="MATLGHTEPPYAGKPEFPMDTIASIMIFLTALATFVILPG

IRGKRLFWLRYVTSIFGAALIGTPVQOLNETINYNSEFTMLGNYAEYAKLE

KGLPDPVLYLAERTPRSPCGIYRQYRLAGHYTSAMLTVAFLCMLANVMSPLYLV

GGYMLLATGIFOLLALFFSMATSLTSPCLHLGASVHTHGPAPMTTLTGILCVL

LGLMAVAHRMQPHRLKAFNQSVDEDPMEWSPEEGILSPRYRSMADSPKSDIPL

SEASSTKAYCKEAKHPKDPDCAL"

ORIGIN

Query Match

27.2%; Score 392; DB 9; Length 2684;

Best Local Similarity 100.0%; Pred. No. 1.2e-205;

Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1027 GGGCTCCTGTCCTCTTCTCCGAGAGGGCCGTGTGAGTCTCCAGTATGTTCCGCCACAGC 1086
 Db 2343 GGGCTCCTGTCCTCTTCTCCGAGAGGGCCGTGTGAGTCTCCAGTATGTTCCGCCACAGC 2284
 QY 1087 GCTCTTCGACCCCTTCTGACCAAGCCGCAAGAGACTGAGCCAGAGAGAGAGGGGCTCA 1146
 Db 2283 GCTCTTCGACCCCTTCTGACCAAGCCGCAAGAGACTGAGCCAGAGAGAGAGGGGCTCA 2224
 QY 1147 CCTCTATCTCGCGGACCACTGACCAAGAGAGAGGGCCCTCTCCAGACTTAAATGATC 1206
 Db 2223 CCTCTATCTCGCGGACCACTGACCAAGAGAGAGGGCCCTCTCCAGACTTAAATGATC 2164
 QY 1207 ACCACTAACCCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCCTTGGAGCATCGCAGG 1266
 Db 2163 ACCACTAACCCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCCTTGGAGCATCGCAGG 2104
 QY 1267 CCGGGAAGCAGTGCCTCCGCGGAGGCTGGGCGGAGAGAGAGCTCCAGAAAGGCACTGAGCGCT 1326
 Db 2103 CCGGGAAGCAGTGCCTCCGCGGAGGCTGGGCGGAGAGAGAGCTCCAGAAAGGCACTGAGCGCT 2044
 QY 1327 GCTGGCGGAGGCTCGGACATCCGAGGAGCCAGGAAAGTCTCTGGGGCGATCTGTA 1386
 Db 2043 GCTGGCGGAGGCTCGGACATCCGAGGAGCCAGGAAAGTCTCTGGGGCGATCTGTA 1984
 QY 1387 AATAAACCTTTTCTTTTCTTTTAAAAA 1418
 Db 1983 AATAAACCTTTTCTTTTCTTTTAAAAA 1952

RESULT 8

AC009700

LOCUS

DEFINITION

SEQUENCE

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AC009700 156534 bp DNA linear HTG 20-APR-2000
 Homo sapiens chromosome 15 clone RP11-163P10 map 15, WORKING DRAFT

SEQUENCE, 15 unordered pieces.

AC009700 AC009700.4 GI:7622346

HTG; HTGS_PHASE1; HTGS_DRAFT.

Homo sapiens (human)

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 156534)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 15, clone RP11-163P10
 Unpublished
 2 (bases 1 to 156534)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, J., Barna, N., Beckert, R., Benn, J., Brown, A., Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Depayre, E., Devon, K., Dewar, K., Doneelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G., Hages, B., Heathford, A., Horton, L., Howland, J. C., Jones, C., Kann, L., Karatas, A., Lenoczky, J., Liu, C., Locke, K., Macdonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Toruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.
 Direct Submission
 Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 20, 2000 this sequence version replaced gi:6056272.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L2336

Center clone name: 163_P_10

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 108837 bases at least Q40

Consensus quality: 130230 bases at least Q30

Consensus quality: 145227 bases at least Q20

Insert size: 157000; agarose-fp

Insert size: 155134; sum-of-contrigs

Quality coverage: 3.2 in Q20 bases; agarose-fp

Quality coverage: 3.3 in Q20 bases; sum-of-contrigs

NOTE

This is a 'working draft' sequence. It currently consists of 15 contrigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contrigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1169: contrig of 1169 bp in length
 * 1170 1269: gap of 100 bp
 * 1270 3335: contrig of 2066 bp in length
 * 3336 3435: gap of 100 bp
 * 3436 5698: contrig of 2263 bp in length
 * 5699 5798: gap of 100 bp
 * 5799 9216: contrig of 3418 bp in length
 * 9217 9316: gap of 100 bp
 * 9317 15633: contrig of 6317 bp in length
 * 15634 15733: gap of 100 bp
 * 15734 21746: contrig of 6013 bp in length
 * 21747 21846: gap of 100 bp
 * 21847 30351: contrig of 8505 bp in length
 * 30352 30451: gap of 100 bp
 * 30452 36509: contrig of 6058 bp in length
 * 36510 36609: gap of 100 bp


```

* 36610 45279: contig of 8670 bp in length
* 45280 45379: gap of 100 bp
* 45380 56851: contig of 11472 bp in length
* 56852 56951: gap of 100 bp
* 56952 67374: contig of 10423 bp in length
* 67375 67474: gap of 100 bp
* 67475 79534: contig of 12060 bp in length
* 79535 79634: gap of 100 bp
* 79635 95432: contig of 15798 bp in length
* 95433 95532: gap of 100 bp
* 95533 120801: contig of 25269 bp in length
* 120802 120901: gap of 100 bp
* 120902 156534: contig of 35633 bp in length.
FEATURES
source
1..156534
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-163P10"
/clone_1b="RPC1-11 Human Male BAC"
1..1169
/note="assembly_fragment"
1270..3335
/note="assembly_fragment"
3436..5698
/note="assembly_fragment"
5799..9216
/note="assembly_fragment"
9317..15633
/note="assembly_fragment"
15734..21746
/note="assembly_fragment"
21847..30351
/note="assembly_fragment"
30452..36509
/note="assembly_fragment"
/note="assembly_fragment
vector_side:left"
36610..45279
/note="assembly_fragment"
45380..56851
/note="assembly_fragment"
56952..67374
/note="assembly_fragment"
67475..79534
/note="assembly_fragment"
79635..95432
/note="assembly_fragment"
95533..120801
/note="assembly_fragment"
120902..156534
/note="assembly_fragment"
ORIGIN
Query Match 27.2%; Score 392; DB 2; Length 156534;
Best Local Similarity 100.0%; Pred. No. 1.2e-205;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1027 GGGCTCCTGCTCTCTCTCGAGGGGCGCTGTAGTCTCCAGTATGTTGGGCCAGC 1086
Db 96131 GGGCTCCTGCTCTCTCTCGAGGGGCGCTGTAGTCTCCAGTATGTTGGGCCAGC 96190
QY 1087 GCTCTTGCACCCCTTCTGACCAAGCGCAAGACTGCAGCCAGAGAGAGGGGGCTCA 1146
Db 96191 GCTCTTGCACCCCTTCTGACCAAGCGCAAGACTGCAGCCAGAGAGAGGGGGCTCA 96250
QY 1147 CCTCTTATCTCGGCGACCACTGCACAGCAGCGCGCTCTCCAGACTTAATAATGTATC 1206
Db 96251 CCTCTTATCTCGGCGACCACTGCACAGCAGCGCGCTCTCCAGACTTAATAATGTATC 96310
QY 1207 ACCACTAACCTGTGAGGGGAGACCAATCTGACTCTTCCCGCCTTGGAGACATCGCAGG 1266

```

```

|||||
Db 96311 ACCACTAACCTGTGAGGGGAGACCAATCTGACTCTTCCCGCCTTGGAGACATCGCAGG 96370
QY 1267 CCGGAGACAGTGCCTGCCAGGCGCTGGCCAGAGAGAGCTCCAGAGAGGGCACTGAGCGCT 1326
Db 96371 CCGGAGACAGTGCCTGCCAGGCGCTGGCCAGAGAGAGCTCCAGAGAGGGCACTGAGCGCT 96430
QY 1327 GCTGGCGGAGGCGCTCGACATCCGACAGGACACGAGAAAGTCTCTGGGGCATCTGTA 1386
Db 96431 GCTGGCGGAGGCGCTCGACATCCGACAGGACACGAGAAAGTCTCTGGGGCATCTGTA 96490
QY 1387 AATAACCTTTTCTTTGTTTAAATAA 1418
Db 96491 AATAACCTTTTCTTTGTTTAAATAA 96522

```

RESULT 9

AC087790

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

```

AC087790 171444 bp DNA linear HTG 09-MAY-2001
Homo sapiens chromosome 15 clone RP11-276K9 map 15, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
AC087790
AC087790.2 GI:13357344
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILLTOP.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Barren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A.,
Camara, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Karatas, A., Laroque, K., Lamazares, R., Landers, T.,
Lehoczky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,
McQuarrie, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M.,
Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Sougen, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2001 this sequence version replaced gi:12408501.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12451
Center clone name: 276_K_9
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

```



```
Consensus quality: 165906 bases at least Q40
Consensus quality: 168587 bases at least Q30
Consensus quality: 169557 bases at least Q20
Insert size: 180000; agarose-fp
Insert size: 170244; sum-of-ctnigs
Quality coverage: 5.7 in Q20 bases; agarose-fp
Quality coverage: 6.0 in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 6091: contig of 6091 bp in length
* 6092 6191: gap of 100 bp
* 6192 7986: contig of 1795 bp in length
* 7987 8086: gap of 100 bp
* 8087 43939: contig of 35853 bp in length
* 43940 44039: gap of 100 bp
* 44040 45298: contig of 1259 bp in length
* 45299 45398: gap of 100 bp
* 45399 47595: contig of 2197 bp in length
* 47596 47695: gap of 100 bp
* 47696 52385: contig of 4690 bp in length
* 52386 52485: gap of 100 bp
* 52486 56310: contig of 3825 bp in length
* 56311 56410: gap of 100 bp
* 56411 71313: contig of 14903 bp in length
* 71314 71413: gap of 100 bp
* 71414 90993: contig of 19580 bp in length
* 90994 91093: gap of 100 bp
* 91094 111459: contig of 20366 bp in length
* 111460 111559: gap of 100 bp
* 11560 137726: contig of 26167 bp in length
* 137727 137826: gap of 100 bp
* 137827 170053: contig of 32227 bp in length
* 170054 170153: gap of 100 bp
* 170154 171444: contig of 1291 bp in length.
Location/Qualifiers
FEATURES
Source
1. 171444
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-276K9"
/clone_lib="RPC1-11 Human Male BAC"
1. 6091
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
misc_feature
6192..7986
/note="assembly_fragment"
8087..43939
/note="assembly_fragment"
44040..45298
/note="assembly_fragment"
45399..47595
/note="assembly_fragment"
47696..52385
/note="assembly_fragment"
52486..56310
/note="assembly_fragment"
56411..71313
/note="assembly_fragment"
71414..90993
/note="assembly_fragment"
91094..111459
/note="assembly_fragment"
111560..137726
misc_feature
```

```
misc_feature /note="assembly_fragment"
137827..170053
/note="assembly_fragment"
misc_feature 170154..171444
/note="assembly_fragment"
clone_end:T7
vector_side:right"
ORIGIN
Query Match 27.2%; Score 392; DB 2; Length 171444;
Best Local Similarity 100.0%; Pred. No. 1.2e-205;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1027 GGGCTTCCTGCTCTTCTCTCGAGGGCCGTGTGAGTCTCCAGTATGTTCCGCCAGC 1086
DB 143780 GGGCTTCCTGCTCTTCTCTCGAGGGCCGTGTGAGTCTCCAGTATGTTCCGCCAGC 143839
QY 1087 GCTCTTCGACCCCTTCTGACCAAGCCCAAGACTGACGCCAGAGAGAGAGGGGCTCA 1146
DB 143840 GCTCTTCGACCCCTTCTGACCAAGCCCAAGACTGACGCCAGAGAGAGAGGGGCTCA 143899
QY 1147 CCTCTTATCTCGGCGACCCCACTGCAAGAGAGGGCCCTCTCCAGACTTAATGATATC 1206
DB 143900 CCTCTTATCTCGGCGACCCCACTGCAAGAGAGGGCCCTCTCCAGACTTAATGATATC 143959
QY 1207 ACCACTAAGCTGTAGGGGAGACCAATCTGACTCTTCCCGCCTTGAGACATCGCAGG 1266
DB 143960 ACCACTAAGCTGTAGGGGAGACCAATCTGACTCTTCCCGCCTTGAGACATCGCAGG 144019
QY 1267 CCGGAAGCACTGCCCGCAGGCGCTGGCCAGAGAGCTCCAGAAAGGACATGACGCT 1326
DB 144020 CCGGAAGCACTGCCCGCAGGCGCTGGCCAGAGAGCTCCAGAAAGGACATGACGCT 144079
QY 1327 GCTGCGGAGGCGCTCGACATCCGAGGACACGAGAAAGTCTCTGGGCGATCTGTA 1386
DB 144080 GCTGCGGAGGCGCTCGACATCCGAGGACACGAGAAAGTCTCTGGGCGATCTGTA 144139
QY 1387 AATAAACCTTTTCTTTTCTTTGTTTAAAAA 1418
DB 144140 AATAAACCTTTTCTTTTCTTTGTTTAAAAA 144171
RESULT 10
AC009700 156534 bp DNA linear HTG 20-APR-2000
LOCUS AC009700/c
DEFINITION Homo sapiens chromosome 15 clone RP11-163P10 map 15, WORKING DRAFT
ACCESSION AC009700 GI:7622346
VERSION AC009700.4
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 156534)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 15, clone RP11-163P10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 156534)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckertly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewart,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marguis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Nioff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tefaye,S., Tornuella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
```

TITLE
JOURNAL

COMMENT

Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

Direct Submission

Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 20, 2000 this sequence version replaced gi:6056272.
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L2336

Center clone name: 163 P_10

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 108837 bases at least Q40

Consensus quality: 130230 bases at least Q30

Consensus quality: 145227 bases at least Q20

Insert size: 157000; agarose-fp

Insert size: 155134; sum-of-coverage

Quality coverage: 3.2 in Q20 bases; agarose-fp

Quality coverage: 3.3 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

* 1 1169: contig of 1169 bp in length
* 1170 1269: gap of 100 bp
* 1270 3335: contig of 2066 bp in length
* 3336 3435: gap of 100 bp
* 3436 5698: contig of 2263 bp in length
* 5699 5798: gap of 100 bp
* 5799 9216: contig of 3418 bp in length
* 9217 9316: gap of 100 bp
* 9317 15633: contig of 6317 bp in length
* 15634 15733: gap of 100 bp
* 15734 21746: contig of 6013 bp in length
* 21747 21846: gap of 100 bp
* 21847 30351: contig of 8505 bp in length
* 30352 30451: gap of 100 bp
* 30452 36509: contig of 6058 bp in length
* 36510 36609: gap of 100 bp
* 36610 45279: contig of 8670 bp in length
* 45280 45379: gap of 100 bp
* 45380 56851: contig of 11472 bp in length
* 56852 56951: gap of 100 bp
* 56952 67374: contig of 10423 bp in length
* 67375 67474: gap of 100 bp
* 67475 79534: contig of 12060 bp in length
* 79535 79634: gap of 100 bp
* 79635 95432: contig of 15798 bp in length
* 95433 95532: gap of 100 bp
* 95533 120801: contig of 25269 bp in length
* 120802 120901: gap of 100 bp
* 120902 156534: contig of 35633 bp in length.

```

FEATURES

source

```

1. 156534
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-163P10"

```

```

misc_feature /clone_1b="RP11-11 Human Male BAC"
1. 1169 /note="assembly_fragment"
misc_feature 1270. 3335 /note="assembly_fragment"
misc_feature 3436. 5698 /note="assembly_fragment"
misc_feature 5799. 9216 /note="assembly_fragment"
misc_feature 9317. 15633 /note="assembly_fragment"
misc_feature 15734. 21746 /note="assembly_fragment"
misc_feature 21847. 30351 /note="assembly_fragment"
misc_feature 30452. 36509 /note="assembly_fragment"
misc_feature /note="assembly_fragment"
clone_end:17
vector_side:left"
36610. 45279 /note="assembly_fragment"
45380. 56851 /note="assembly_fragment"
56952. 67374 /note="assembly_fragment"
67475. 79534 /note="assembly_fragment"
79635. 95432 /note="assembly_fragment"
95533. 120801 /note="assembly_fragment"
120902. 156534 /note="assembly_fragment"

```

ORIGIN

```

Query Match 24.6%; Score 354; DB 2; Length 156534;
Best Local Similarity 99.8%; Pred. No. 1.4e-184;
Matches 404; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAGTAACGGCTACAGACAGTGAAGTAAGTTGCTCGCCGCTAGAAAAAAGCTGTGCG 60
Db 7148 AAAGTAACGGCTACAGACAGTGAAGTAAGTTGCTCGCCGCTAGAAAAAAGCTGTGCG 7089

QY 61 GTACCAACCCAGAGCGTTGAGAGAGAGAGCCCACTCCAGCTTCTTAACGAGAGGTGCA 120
Db 7088 GTACCAACCCAGAGCGTTGAGAGAGAGAGCCCACTCCAGCTTCTTAACGAGAGGTGCA 7029

QY 121 GAACTCAGACTTCACCAAGCCCACTGGTCCAGCCTTGTAAGCAAGAGAGAGAGAGAC 180
Db 7028 GAACTCAGACTTCACCAAGCCCACTGGTCCAGCCTTGTAAGCAAGAGAGAGAGAGAC 6969

QY 181 GCGCTCTCCGCGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Db 6968 GCGCTCTCCGCGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6909

QY 241 ACTGAGCCGCGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Db 6908 ACTGAGCCGCGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6849

QY 301 CCGCATGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 6848 CCGCATGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6789

QY 361 GCAGCAAGCTTCTGCTCATCTTGGCCGGGAGATCCGTGGCACTCG 405
Db 6788 GCAGCAAGCTTCTGCTCATCTTGGCCGGGAGATCCGTGGCACTCG 6744

```

RESULT 11

```

HS183H12F/c 217 bp DNA linear PRI 18-OCT-1995
LOCUS HS183H12F/c H.sapiens Cpg island DNA genomic MseI fragment, clone 183h12,
DEFINITION forward read cpj183h12.ft1a.

```

ACCESSION 257552
VERSION 257552.1 GI:1028783
KEYWORDS Cpg island; genomic MseI fragment.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
TITLE Purification of Cpg islands using a methylated DNA binding column
JOURNAL Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE 94282070
PUBMED 8012384
REFERENCE 2 (bases 1 to 217)
AUTHORS Dodsworth,S.J., Huckle,E., Wilkinson,P. and Micklem,G.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk
COMMENT Vector: pGEM-5Zf(-)
Clones are available from the UK MRC Human Genome Mapping Project
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
http://www.hgmp.mrc.ac.uk/ for details
or contact: biohelp@hgmp.mrc.ac.uk.
FEATURES
source 1..217
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="183h12"
/sex="male"
/tissue_type="blood"
/clone_lib="CGI-1"
/dev_stage="adult"

ORIGIN

Query Match 15.1%; Score 217; DB 9; Length 217;
Best Local Similarity 100.0%; Pred. No. 1.4e-108;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1198 AAATGTATCACCCTAACCCTGTGAGGGGGACCCCAATCTGACTCCTTCCCGCCTTGGGA 1257
DB 217 AAATGTATCACCCTAACCCTGTGAGGGGGACCCCAATCTGACTCCTTCCCGCCTTGGGA 158
QY 1258 CATCGCAGCGCGGGAAGACAGTGCCTCCGCAAGGCTCCAGGAAGGCA 1317
DB 157 CATCGCAGCGCGGGAAGACAGTGCCTCCGCAAGGCTCCAGGAAGGCA 98
QY 1318 CTGAGCGCTGCTGGCGCGGAGCCTCGGACATCCGACAGGACCAAGGAAAGTCTCTGGGG 1377
DB 97 CTGAGCGCTGCTGGCGCGGAGCCTCGGACATCCGACAGGACCAAGGAAAGTCTCTGGGG 38
QY 1378 CGATCTGTAATAACCTTTTCTTTCTTTTCTTTT 1414
DB 37 CGATCTGTAATAACCTTTTCTTTCTTTTCTTTT 1

RESULT 12
BC029819/c 1923 bp tRNA linear PRI 06-OCT-2003
LOCUS Homo sapiens hypothetical protein FLJ32334, mRNA (cDNA clone
DEFINITION MGC:35310 IMAGE:5177619), complete cds.
ACCESSION BC029819
VERSION BC029819
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1923)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
REMARK

NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunatne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegel, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Navaratil,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRAK Plate: 51 Row: 0 Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21389332.
FEATURES
source 1..1923
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:35310 IMAGE:5177619"
/tissue_type="Brain, Lung, Testis, adult, pooled whole"
/clone_lib="NIH MGC_115"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..1923
/gene="FLJ32334"
/db_xref="LOCUSID:90527"
401..1852
/codon_start=1
/product="FLJ32334 protein"
/protein_id="AAH29819.1"
/db_xref="GI:20987584"
/db_xref="LOCUSID:90527"

gene
CDS
Diacchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramsen,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.T., Skalska,U., Smolius,D.E.,
Schnierch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1923)
Strausberg,R.
Direct Submission
Submitted (06-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

ORIGIN DLPKCDYRRDARAFHLEPTCVTSRHTPLI"

Query Match 11.5%; Score 166; DB 9; Length 1923;
Best Local Similarity 100.0%; Pred. No. 2.7e-80;
Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1216 CTGTGAGGGGAGACCAATCTGACTCTTCCCGCCTTGGACATCGACGCGGGAAGC 1275
|||||
Db 1557 CTGTGAGGGGAGACCAATCTGACTCTTCCCGCCTTGGACATCGACGCGGGAAGC 1498
|||||

QY 1276 AGTGCCCGCCAGAGCCTGGCCAGAGAGCTCCAGAAAGGCACTGAGCGCTGCGCGG 1335
|||||
Db 1497 AGTGCCCGCCAGAGCCTGGCCAGAGAGCTCCAGAAAGGCACTGAGCGCTGCGCGG 1438
|||||

QY 1336 AGGCCTCGACATCCGACAGCAGGAAAGTCTCTGGGCGCAT 1381
|||||
Db 1437 AGGCCTCGACATCCGACAGCAGGAAAGTCTCTGGGCGCAT 1392
|||||

RESULT 13
HS183A12R 218 bp DNA linear PRI 19-OCT-1995
LOCUS HS183A12R
DEFINITION H.sapiens Cpg island DNA genomic MseI fragment, clone 183a12,
reverse read cpg183a12.r1a.
ACCESSION Z59954
VERSION Z59954.1 GI:1031867
KEYWORDS Cpg island; genomic MseI fragment.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.
TITLE Purification of Cpg islands using a methylated DNA binding column
JOURNAL Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE 94282070
PUBMED 8012384

REFERENCE
2 (bases 1 to 218)
Dodsworth, S.J., Huckle, E., Wilkinson, P. and Micklem, G.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk
Vector: pGEM-5Zf(-)
Clones are available from the UK MRC Human Genome Mapping Project
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
http://www.hgmp.mrc.ac.uk/ for details
or contact: biohelp@hgmp.mrc.ac.uk.
Location/Qualifiers

FEATURES
source 1. 218
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="183a12"
/sex="male"
/tissue_type="blood"
/clone_lib="CGI-1"
/dev_stage="adult"

ORIGIN

Query Match 9.4%; Score 136; DB 9; Length 218;
Best Local Similarity 100.0%; Pred. No. 1.2e-63;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1268 CGGGAAGCAGTCCCGCCAGGCTGGCCAGAGAGCTCCAGAAAGGCACTGACGCTG 1327
|||||
Db 72 CGGGAAGCAGTCCCGCCAGGCTGGCCAGAGAGCTCCAGAAAGGCACTGACGCTG 131
|||||

QY 1328 CTGGCGGAGGCTCGGACATCCGACAGCAGGAAAGTCTCTGGGCGCATCTGTAA 1387
|||||
Db 132 CTGGCGGAGGCTCGGACATCCGACAGCAGGAAAGTCTCTCTGGGCGCATCTGTAA 191
|||||

QY 1388 ATAAACCTTTTCT 1403

Db 192 ATAAACCTTTTCT 207
|||||

RESULT 14
BC004161 1725 bp mRNA linear PRI 12-NOV-2003
LOCUS BC004161
DEFINITION Homo sapiens transmembrane 4 superfamily member tetraspan NET-7,
mRNA (cDNA clone MGC:2447 IMAGE:2958221), complete cds.
ACCESSION BC004161
VERSION BC004161.2 GI:33872773
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 1725)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettelman, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallus, D.E.,
Scherer, A., Schein, J.E., Jones, S.J. and Marra, M.A.
TITLE Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 12477932
PUBMED 12477932

REFERENCE
2 (bases 1 to 1725)
Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
NIH-MGC Project URL: http://mgc.nci.nih.gov
On Aug 19, 2003 this sequence version replaced gi:13278776.
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

FEATURES
source 1. 1725
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:2447 IMAGE:2958221"
/tissue_type="Kidney, renal cell adenocarcinoma"
/clone_lib="NIH_MGC_14"

```
/lab_host="DH10B-R"
/notes="vector: potB7"
1. .1725
/gene="NET-7"
/db_xref="LocustID:23555"
120. .1004
/codon_start=1
/product="tetraspan NET-7"
/protein_id="AAH04161.1"
/db_xref="GI:13278777"
/db_xref="LocustID:23555"
/translation="MPRGDSQVRYCARFSYLWLFSLIYSTVFWLIGALVLSVGIY
AEVERQKYKTLLESAPLAPAILLLGVVMFVSPFGLASLRDNLVLLQAFMYLIGIC
LIMELIGVVALTFRNQOTIDFLNDNRGIENYDDLDKFNIMDFVQKKFCGGEDY
RDWSKNQYHDCSAPGLACGVPYTCIRNTTEVNTMGYKTIIDKERSVQDVIYVRG
CTNAVLIWMFMDNTIMAGILGLLPQFLGVLTLTYTRVEDIMEHSVTDLGLPG
AKPSVEAGTGCCLCYPN"
174. .854
/notes="transmembrane4; Region: Tetraspanin family"
/db_xref="CDD:pfam00335"

misc_feature
174. .854

ORIGIN
Query Match      2.7%; Score 39; DB 9; Length 1725;
Best Local Similarity 100.0%; Pred. No. 7.8e-10;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1403 TTTTGTTTTAAAAA
Db      1683 TTTTGTTTTAAAAA

RESULT 15
BC003157      1726 bp      mRNA      linear      PRI 03-OCT-2003
LOCUS      Homo sapiens transmembrane 4 superfamily member tetraspan NET-7,
DEFINITION      mRNA (cDNA clone MGC:4120 IMAGE:2958221), complete cds.
ACCESSION      BC003157
VERSION      BC003157.1 GI:13111968
KEYWORDS
SOURCE      MGC.
ORGANISM      Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1726)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Datchenko,L., Marusik,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,U.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahy,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Boutfard,G.G., Blakesley,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,V.S., Krzywinski,M.I., Skalska,U., Smalins,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
JOURNAL
MEDLINE
PUBMED      12477932
2 (bases 1 to 1726)
Strausberg,R.
Direct Submission
Submitted (13-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
```

```
REMARK      USA
COMMENT      NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalins, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRAL Plate: 9 Row: n Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21264576.

FEATURES
source
1. .1726
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:4120 IMAGE:2958221"
/tissue_type="Kidney, renal cell adenocarcinoma"
/clone_id="DH10B-R"
/lab_host="NIH MGC_14"
/notes="vector: potB7"
1. .1726
/gene="NET-7"
/db_xref="LocustID:23555"
120. .1004
/codon_start=1
/product="tetraspan NET-7"
/protein_id="AAH03157.1"
/db_xref="GI:13111969"
/db_xref="LocustID:23555"
/translation="MPRGDSQVRYCARFSYLWLFSLIYSTVFWLIGALVLSVGIY
AEVERQKYKTLLESAPLAPAILLLGVVMFVSPFGLASLRDNLVLLQAFMYLIGIC
LIMELIGVVALTFRNQOTIDFLNDNRGIENYDDLDKFNIMDFVQKKFCGGEDY
RDWSKNQYHDCSAPGLACGVPYTCIRNTTEVNTMGYKTIIDKERSVQDVIYVRG
CTNAVLIWMFMDNTIMAGILGLLPQFLGVLTLTYTRVEDIMEHSVTDLGLPG
AKPSVEAGTGCCLCYPN"
180. .896
/notes="transmembrane4; Region: Tetraspanin family"
/db_xref="CDD:pfam00335"

misc_feature
180. .896

ORIGIN
Query Match      2.7%; Score 39; DB 9; Length 1726;
Best Local Similarity 100.0%; Pred. No. 7.8e-10;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1403 TTTTGTTTTAAAAA
Db      1683 TTTTGTTTTAAAAA
```

Search completed: February 24, 2004, 01:57:04
Job time : 5689.97 secs

CC of tumour cells and in vaccines for prophylactic and therapeutic
CC treatment of cancers, particularly ovarian or colon cancer, autoimmune
CC diseases and related conditions

Sequence 1441 BP; 289 A; 466 C; 392 G; 294 T; 0 U; 0 Other;

Query Match	100.0%;	Score 1441;	DB 3;	Length 1441;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1441; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	1	AAAGTAA	CGGCTAC	GACAG	AGTGA	AAATAG	TTCCGCTCG	CGCCGGCTAG	AAAAAACTCTGTCG	60
Db	1	AAAGTAA	CGGCTAC	GACAG	AGTGA	AAATAG	TTCCGCTCG	CGCCGGCTAG	AAAAAACTCTGTCG	60
QY	61	GTACCA	CCCCCAG	AGCGTTG	AGAGCAG	CCCCCACTCC	CAGCTTCC	TTAACGAGAGG	AGCA	120
Db	61	GTACCA	CCCCCAG	AGCGTTG	AGAGCAG	CCCCCACTCC	CAGCTTCC	TTAACGAGAGG	AGCA	120
QY	121	GGACTCAG	ACTTAC	CAGAGCC	CACTCGTCC	AGCCTTGTAC	GCMAAGAGAC	GCACAGAG	AGC	180
Db	121	GGACTCAG	ACTTAC	CAGAGCC	CACTCGTCC	AGCCTTGTAC	GCMAAGAGAGC	GCACAGAG	AGC	180
QY	181	GCGCTC	TCGCCG	CGTCCAG	GAGCCCG	AGCTTGCTG	CCCTG	CCGCGCTG	CGTCAGC	240
Db	181	GCGCTC	TCGCCG	CGTCCAG	GAGCCCG	AGCTTGCTG	CCCTG	CCGCGCTG	CGTCAGC	240
QY	241	ACTCGG	CCGGCG	TGCAG	CATGAC	CCCTGTGAA	CGGCGTACTG	CCCTTTTAA	CCCCCAG	300
Db	241	ACTCGG	CCGGCG	TGCAG	CATGAC	CCCTGTGAA	CGGCGTACTG	CCCTTTTAA	CCCCCAG	300
QY	301	CGGATG	CCGCGCAG	GCCTTCA	GGCCTTCC	ACTGCTCAT	CGTAA	TCTAGT	TTTTGGCTCTA	360
Db	301	CGGATG	CCGCGCAG	GCCTTCA	GGCCTTCC	ACTGCTCAT	CGTAA	TCTAGT	TTTTGGCTCTA	360
QY	361	GCAGCAG	CTTCTG	CTCATCTT	TGCGGGGAT	CCGTG	GCCACTG	CGCGTGT	TTGGTTG	420
Db	361	GCAGCAG	CTTCTG	CTCATCTT	TGCGGGGAT	CCGTG	GCCACTG	CGCGTGT	TTGGTTG	420
QY	421	GTGAGAG	TTCTTCTCAG	TCTGTTATAG	GGCGGAATTG	TGGCTGTG	CACTTCA	GTGCA	TTCA	480
Db	421	GTGAGAG	TTCTTCTCAG	TCTGTTATAG	GGCGGAATTG	TGGCTGTG	CACTTCA	GTGCA	TTCA	480
QY	481	GAATG	TTCTG	TGGTAC	GTAAACCA	CAACATCTT	CAAA	AGCCTT	CAGCGAGCGCG	540
Db	481	GAATG	TTCTG	TGGTAC	GTAAACCA	CAACATCTT	CAAA	AGCCTT	CAGCGAGCGCG	540
QY	541	GTTACAG	CCCGTGT	CGGTCTG	CTGCTGG	CGCTGAGAGG	GCATTATAT	TACACT	CACAGAGG	600
Db	541	GTTACAG	CCCGTGT	CGGTCTG	CTGCTGG	CGCTGAGAGG	GCATTATAT	TACACT	CACAGAGG	600
QY	601	ACCCAG	TGCATCAG	CTGAACG	AGCATTG	ACTAC	CAACGAG	CAGTTCA	CTGCGCTG	660
Db	601	ACCCAG	TGCATCAG	CTGAACG	AGCATTG	ACTAC	CAACGAG	CAGTTCA	CTGCGCTG	660
QY	661	AAAGAG	AAATTAC	CGCCGGGAG	TACGCGGAAC	GCAC	TGGAGAA	GGGGCTG	CCCGAACCCAGTG	720
Db	661	AAAGAG	AAATTAC	CGCCGGGAG	TACGCGGAAC	GCAC	TGGAGAA	GGGGCTG	CCCGAACCCAGTG	720
QY	721	CTTAC	CTGGCGGAG	AGATTAC	ACCCAGTAG	CCCTTG	CGGCTGTAC	CAACAG	TACAC	780
Db	721	CTTAC	CTGGCGGAG	AGATTAC	ACCCAGTAG	CCCTTG	CGGCTGTAC	CAACAG	TACAC	780
QY	781	CTGGCGG	ACACTAC	GCCTCG	GCACGCTAT	GGGTGG	CGTTCTG	GGCTCCT	CTCC	840
Db	781	CTGGCGG	ACACTAC	GCCTCG	GCACGCTAT	GGGTGG	CGTTCTG	GGCTCCT	CTCC	840
QY	841	AACGTG	CTGCTT	CCAGCG	CGCCCGCTT	ACGAGG	CGCTTG	GCAC	CCGGA	900
Db	841	AACGTG	CTGCTT	CCAGCG	CGCCCGCTT	ACGAGG	CGCTTG	GCAC	CCGGA	900
QY	901	GCGTTG	CGGCTCTT	CGGGTCTT	GGCTTG	GCCTTCA	TCTTAG	CGTGC	GTGCGGCTG	960
Db	901	GCGTTG	CGGCTCTT	CGGGTCTT	GGCTTG	GCCTTCA	TCTTAG	CGTGC	GTGCGGCTG	960

QY	961	CTCCGCTAGGCTCTCCGGGCTCACCACCTAGTACGGCGCCCTTCTGGGTACGCTG	1020
Db	961	CTCCGCTAGGCTCTCCGGGCTCACCACCTAGTACGGCGCCCTTCTGGGTACGCTG	1020
QY	1021	GCACCCGGGCTCTGTGCTCTTCTCGAGAGGGCGGTGTGAGTCTCCAGTATGTTCCG	1080
Db	1021	GCACCCGGGCTCTGTGCTCTTCTCGAGAGGGCGGTGTGAGTCTCCAGTATGTTCCG	1080
QY	1081	CCGAGCGCTCTTCGACCCCTTCTGACCAAGCGCCAAGACTCGACGAGAGAGAGG	1140
Db	1081	CCGAGCGCTCTTCGACCCCTTCTGACCAAGCGCCAAGACTCGACGAGAGAGAGG	1140
QY	1141	GGCTCACCTCTTATCCTCGGCGACCCCACTGCACAAAGCAGGCGCTCTCCAGACTTAAAA	1200
Db	1141	GGCTCACCTCTTATCCTCGGCGACCCCACTGCACAAAGCAGGCGCTCTCCAGACTTAAAA	1200
QY	1201	TGTATCACCACTAACTGTGAGGGGGACCCAACTGTGACTCTTCCCGCTTGGACAT	1260
Db	1201	TGTATCACCACTAACTGTGAGGGGGACCCAACTGTGACTCTTCCCGCTTGGACAT	1260
QY	1261	CGCAGGCGCGGAGCAGTGTCCCGCAGGCTTGGGCCAGGAGAGCTCCAGGAGGGCACTG	1320
Db	1261	CGCAGGCGCGGAGCAGTGTCCCGCAGGCTTGGGCCAGGAGAGCTCCAGGAGGGCACTG	1320
QY	1321	AGCGCTGTGGCGCGAGGCTTCGACATTCGACAGGACCAAGGGAAGTCTCTGGGGCGA	1380
Db	1321	AGCGCTGTGGCGCGAGGCTTCGACATTCGACAGGACCAAGGGAAGTCTCTGGGGCGA	1380
QY	1381	TCTGTAATAAACCCTTTTCTTTCTTTGTTTTTAAAAA	1440
Db	1381	TCTGTAATAAACCCTTTTCTTTCTTTGTTTTTAAAAA	1440
QY	1441	A 1441	
Db	1441	A 1441	

Accession	Gene	Protein	Location	Qualifiers
RESULT 2				
AAA96505	AAA96505	standard; cDNA; 1421 BP.		
XX	AAA96505;			
AC				
XX	08-FEB-2001	(first entry)		
DT				
XX				
DE		cDNA encoding a human transmembrane protein.		
XX				
KW		Human; transmembrane protein; cell proliferation disorder; myeloma;		
KW		reproductive disorder; smooth muscle disorder; neurological disorder;		
KW		arteriosclerosis; leukaemia; acquired immunodeficiency syndrome; AIDS;		
KW		allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;		
KW		Alzheimer's disease; Tourette's disorder; ss.		
OS				
XX	Homo sapiens.			
XX				
FH	Key		Location/Qualifiers	
FT	CDS		267..1229	
FT			/*tag= a	
XX				
XX				
PN	W0200056891-A2.			
XX				
PD	28-SEP-2000.			
XX				
PF	22-MAR-2000; 2000WO-US007817.			
XX				
XX	22-MAR-1999; 99US-0125537P.			
PR	16-JUN-1999; 99US-0139565P.			
XX				
XX				
PA	(INCY-) INCYTE PHARM INC.			
XX				
PI	Yue H, Lal P, Tang YT, Hillman JL, Reddy R, Bandman O;			
PI	Baughn MR, Lu DM, Azimzai Y, Yang J;			

Yue H, Lal P, Tang YT, Hillman JL, Reddy R, Bardman O, Baughn MR, Lu DM, Azimzai Y, Yang J;

XX WPI; 2000-579485/54.
DR P-PSDB; AAB18992.
XX
PT New human transmembrane proteins are used to treat a disease or condition
PT associated with decreased expression of functional HTMP e.g. Tourette's
PT disorder, angina and leukemia.
XX
PS Claim 4; Page 129; 130pp; English.
XX
CC The present sequence encodes a human transmembrane proteins (HTMP).
CC Agonists and antagonists of the protein are used to treat a disease or
CC condition associated with overexpression of the protein. Diseases and
CC conditions which can be treated include include cell proliferative,
CC immunological, reproductive, smooth muscle and neurological disorders
CC e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency
CC syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,
CC stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The
CC polynucleotides may be used to detect and quantify gene expression in
CC biopsied tissues where protein expression may be correlated with disease
CC e.g. to determine absence, presence or excess expression of HTMP or to
CC monitor regulation of HTMP expression during therapeutic intervention
XX
SQ Sequence 1421 BP; 264 A; 468 C; 395 G; 294 T; 0 U; 0 Other;
Query Match 98.1%; Score 1413; DB 3; Length 1421;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAGTAACGGCTACAGACAGAGAGAAATAGTTTCGCTCGCCGGCTAGAAAACCTGTGCG 60
DB 9 AAAGTAACGGCTACAGACAGAGAGAAATAGTTTCGCTCGCCGGCTAGAAAACCTGTGCG 68
QY 61 GTACCAACCCCGAGAGCGTTGAGAGACAGCCCACTCCACGCTTCTTAAAGAGAGTGCA 120
DB 69 GTACCAACCCCGAGAGCGTTGAGAGACAGCCCACTCCACGCTTCTTAAAGAGAGTGCA 128
QY 121 GGACTCAGACTTCACACAGCCCACTCGGTCCAGCCTTGTAAGCAAGAGACGCCAAGAC 180
DB 129 GGACTCAGACTTCACACAGCCCACTCGGTCCAGCCTTGTAAGCAAGAGACGCCAAGAC 188
QY 181 GCGCTCTCCCGCTCCAGAGCAGCCCGAGCTTGCTGCTGCTGCTGCTGCTGCTGCTG 240
DB 189 GCGCTCTCCCGCTCCAGAGCAGCCCGAGCTTGCTGCTGCTGCTGCTGCTGCTGCTG 248
QY 241 ACTCGCGCGCGGTGAGCATGACCTGTGGAAGCGGCGTACTGCTTTTAAACCCAGGCC 300
DB 249 ACTCGCGCGCGGTGAGCATGACCTGTGGAAGCGGCGTACTGCTTTTAAACCCAGGCC 308
QY 301 CGGCATGCGGAGGCTTCAGCGCTTCCACTGCTCATCGTTATTTAGTGTGCTCTA 360
DB 309 CGGCATGCGGAGGCTTCAGCGCTTCCACTGCTCATCGTTATTTAGTGTGCTCTA 368
QY 361 GCAGCAAGCTTCTGCTCATCTTGCCGGGATCCGTGGCCACTCGCGCTGTTGTTG 420
DB 369 GCAGCAAGCTTCTGCTCATCTTGCCGGGATCCGTGGCCACTCGCGCTGTTGTTG 428
QY 421 GTGAGAGTCTTCTCAGTCTGTTCATAGGCGCAGAAATTGTGCTGTGCACTTCAGTGA 480
DB 429 GTGAGAGTCTTCTCAGTCTGTTCATAGGCGCAGAAATTGTGCTGTGCACTTCAGTGA 488
QY 481 GAATGTTCTGTGGTACAGTGAACACCAACACATCTTACAAGCCTTCAGCGCAGCGCG 540
DB 489 GAATGTTCTGTGGTACAGTGAACACCAACACATCTTACAAGCCTTCAGCGCAGCGCG 548
QY 541 GTTACAGCCCGTGTGCTGCTGCTGCGGCGCTGAGGGGCGATTATATATACACTCAGAGG 600
DB 549 GTTACAGCCCGTGTGCTGCTGCTGCGGCGCTGAGGGGCGATTATATATACACTCAGAGG 608
QY 601 ACCCAGTGCATCAGCTGAAGAGAGCAATTGACTACAAGCAGAGTTCACCTGCGCTG 660
DB 609 ACCCAGTGCATCAGCTGAAGAGAGCAATTGACTACAAGCAGAGTTCACCTGCGCTG 668

QY 661 AAAGAGATTATACCGCGGAGTACGCGAAGCACTGAGAAAGGGGCTGCCGAGCCAGTG 720
DB 669 AAAGAGATTATACCGCGGAGTACGCGAAGCACTGAGAAAGGGGCTGCCGAGCCAGTG 728
QY 721 CTCTACCTGCGGAGAAAGTTACACCGAGTAGCCCTTGCGGCTGTACACCAAGTACAC 780
DB 729 CTCTACCTGCGGAGAAAGTTACACCGAGTAGCCCTTGCGGCTGTACACCAAGTACAC 788
QY 781 CTGGCGGAGACACTACGCTCGGCCACGCTATGGGTGGCGTTCTGCTTCCCTCTCTCC 840
DB 789 CTGGCGGAGACACTACGCTCGGCCACGCTATGGGTGGCGTTCTGCTTCCCTCTCTCC 848
QY 841 AACGTGCTGCTTCCACAGCGCGCGCGCTTACGAGAGGCTGGCACTGTCACCAACCGGA 900
DB 849 AACGTGCTGCTTCCACAGCGCGCGCGCTTACGAGAGGCTGGCACTGTCACCAACCGGA 908
QY 901 GCCTTCGCGCTCTTCCGAGGTCTTCCGCTTCCGCTTCCATCTTACGCTGCGCTGCGCG 960
DB 909 GCCTTCGCGCTCTTCCGAGGTCTTCCGCTTCCGCTTCCATCTTACGCTGCGCTGCGCG 968
QY 961 CTCCGCTTACGCTCTCCGCGCTTCCACACTCAGTACGCGCGCGCTTCTGGGTACGCTG 1020
DB 969 CTCCGCTTACGCTCTCCGCGCTTCCACACTCAGTACGCGCGCGCTTCTGGGTACGCTG 1028
QY 1021 GCAACCGCGCTCTGCTGCTCTTCTCTCGAGAGGCGCGTGTGAGTCTCCAGTATGTTGG 1080
DB 1029 GCAACCGCGCTCTGCTGCTCTTCTCTCGAGAGGCGCGTGTGAGTCTCCAGTATGTTGG 1088
QY 1081 CCCAGCGCTCTTCCGACCCCTTCCGACCAAGCGCCAAAGAGAGTGCAGCCAGAGAGGG 1140
DB 1089 CCCAGCGCTCTTCCGACCCCTTCCGACCAAGCGCCAAAGAGAGTGCAGCCAGAGAGGG 1148
QY 1141 GGCTCACCCTTATCTCGGACCCACTGCACAAAGAGCGCGCTTCCAGACTTAA 1200
DB 1149 GGCTCACCCTTATCTCGGACCCACTGCACAAAGAGCGCGCTTCCAGACTTAA 1208
QY 1201 TGTATCACCACTAATCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGACAT 1260
DB 1209 TGTATCACCACTAATCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGACAT 1268
QY 1261 CGCAGCGCGGAGAGAGTGCAGCGGCTGCGGCGCAGAGAGCTCCAGAAAGGCACTG 1320
DB 1269 CGCAGCGCGGAGAGAGTGCAGCGGCTGCGGCGCAGAGAGCTCCAGAAAGGCACTG 1328
QY 1321 AGCGTGTGCGGCGAGGCTCGGACATCCGAGGACCAAGGAAAGTCTCTGGGGCGA 1380
DB 1329 AGCGTGTGCGGCGAGGCTCGGACATCCGAGGACCAAGGAAAGTCTCTGGGGCGA 1388
QY 1381 TCTGTAATAAACCTTTTCTTTGTTT 1413
DB 1389 TCTGTAATAAACCTTTTCTTTGTTT 1421
RESULT 3
ADD19228
ID ADD19228 standard; cDNA; 1491 BP.
XX
AC ADD19228;
XX
DE Human cDNA from secreted protein gene 45.
XX
KW human secreted protein; cytosolic; antibacterial; virucide;
KW neuroprotective; gynaecological; gastrointestinal-Gen; cardiac;
KW cardiovascular-Gen; nephroretropic; antiinflammatory; muscular-Gen;
KW respiratory-Gen; immunosuppressive; cerebroprotective; vasotropic;
KW nootropic; antiallergic; cancer; bacterial infection; viral infection;
KW neural disorder; immune system disorder; blood disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW inflammatory disorder; proliferative disorder; human; ss; gene.
XX

OS Homo sapiens.
XX WO2003052377-A2.
XX 26-JUN-2003.
PD 06-NOV-2002; 2002WO-US035606.
XX 07-NOV-2001; 2001US-0331046P.
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
PI Rosen CA, Ruben SW;
XX
DR WPI; 2003-533050/50.
DR P-PSDB; ADD19303.
XX
XX
PT New isolated nucleic acids encoding signal transduction pathway component
PT polypeptides, useful for diagnosing, treating, and/or preventing
PT disorders, such as cancer, infections, cardiovascular and inflammatory
PT diseases.
XX
XX
PS Claim 1; SEQ ID NO 55; 554pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (cDNA)
CC encoding a human secreted protein, representing one of 85 novel genes.
CC Also included are recombinant vectors, host cells (expressing the
CC protein), the secreted proteins (including their fragments, epitopes and
CC homologues), an isolated antibody that binds specifically to the protein,
CC diagnosing a pathological condition or susceptibility to a pathological
CC condition (comprising determining the presence or absence of a mutation
CC in the nucleic acid and diagnosing a condition based on the presence or
CC absence of the mutation), diagnosing a pathological condition or
CC susceptibility to a pathological condition (comprising determining the
CC presence or amount of expression of the protein in a biological sample
CC and diagnosing a condition based on the presence or amount of expression
CC of the protein), preventing, treating or ameliorating a medical condition
CC by administering the nucleic acid or protein to a mammalian subject,
CC identifying a binding partner to the protein, the gene corresponding to
CC the cDNA sequence, and identifying an activity in a biological assay
CC (comprising expressing the nucleic acid in a cell, isolating the
CC supernatant, detecting an activity in a biological assay and identifying
CC the protein in the supernatant having the activity). The nucleic acids
CC and proteins display the following activities: Cytostatic, antibacterial,
CC Virucide, Neuroprotective, Gynaecological, Gastrointestinal-Gen,
CC Cardiant, Cardiovascular-Gen, Nephrotropic, Antiinflammatory, Muscular-
CC Gen, Respiratory-Gen, Immunosuppressive, Cerebroprotective, Vasotropic,
CC Nootropic, Antiallergic. The methods and compositions of the present
CC invention are useful for diagnosing, treating, preventing and/or
CC prognosticating disorders related to the novel polypeptides, such as
CC cancer, bacterial or viral infections, and neural, immune system, blood,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal, inflammatory or proliferative disorders (many examples of these
CC diseases and disorders are given in the specification). The present
CC sequence encodes a novel secreted protein of the invention.
XX
SQ Sequence 1491 BP; 293 A; 485 C; 410 G; 302 T; 0 U; 1 Other;
Query Match 92.7%; Score 1336; DB 9; Length 1491;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 161 GCGCTCTCCCGCGTCCAGGACAGCCCGAGCTTGCTGCTTGCTGCGCCGCTGGTGACG 240
DB 224 GCGCTCTCCCGCGTCCAGGACAGCCCGAGCTTGCTGCTTGCTGCGCCGCTGGTGACG 283
QY 241 ACTGCGCCGCGTGACAGATGACCTGTGGAACGGCGTACTGCTTTTACCCGACGCC 300
DB 284 ACTGCGCCGCGTGACAGATGACCTGTGGAACGGCGTACTGCTTTTACCCGACGCC 343
QY 301 CGCATGCGCGAGGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCA 360
DB 344 CGCATGCGCGAGGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCA 403
QY 361 GCAGCAAGCTTCTGCTCATCTTGGCCGGGATCCGTGGCCACTGCGCGTGTGGTGG 420
DB 404 GCAGCAAGCTTCTGCTCATCTTGGCCGGGATCCGTGGCCACTGCGCGTGTGGTGG 463
QY 421 GTGAGAGTTCTTCTCAGTCTGTTCAATAGGCGCAGAAATTGTGCTGTGCACTTCA 480
DB 464 GTGAGAGTTCTTCTCAGTCTGTTCAATAGGCGCAGAAATTGTGCTGTGCACTTCA 523
QY 481 GAATGTTCTGCGGTACAGTGAACACCAACATCTTCAAAAGCTTCAAGCGCGCGC 540
DB 524 GAATGTTCTGCGGTACAGTGAACACCAACATCTTCAAAAGCTTCAAGCGCGCGC 583
QY 541 GTTACAGCCCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
DB 584 GTTACAGCCCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 643
QY 601 ACCCGAGTGCATCAGCTGAACAGAGACCATTTGACTACAACGAGCAGTTCACTGCGCTG 660
DB 644 ACCCGAGTGCATCAGCTGAACAGAGACCATTTGACTACAACGAGCAGTTCACTGCGCTG 703
QY 661 AAGAGATTACGCGCGCGAGTACGCGAAGCAGCAGTGAAGAGGGGCTGCGGACCGAGTG 720
DB 704 AAGAGATTACGCGCGCGAGTACGCGAAGCAGCAGTGAAGAGGGGCTGCGGACCGAGTG 763
QY 721 CTCTAAGTGGGAGAGATTACACACCGAGTACGCGGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 764 CTCTAAGTGGGAGAGATTACACACCGAGTACGCGGCTGCTGCTGCTGCTGCTGCTGCT 823
QY 781 CTGGGCGGACATACGCTGCGGCGCAGCTATGAGTGGGCTGCTGCTGCTGCTGCTGCT 840
DB 824 CTGGGCGGACATACGCTGCGGCGCAGCTATGAGTGGGCTGCTGCTGCTGCTGCTGCT 883
QY 841 AACGTGCTGCTTCCACGCGCGCGCGCTTACGAGGCGCTGGACCTGACCAACCGGA 900
DB 884 AACGTGCTGCTTCCACGCGCGCGCGCTTACGAGGCGCTGGACCTGACCAACCGGA 943
QY 901 GCTTGGCGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTC 960
DB 944 GCTTGGCGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTC 1003
QY 961 CTCCGCTAGGCTCTCTCGCGCTGACCACTCAGTACGCGCGCGCTTTCGCGGCTTTC 1020
DB 1004 CTCCGCTAGGCTCTCTCGCGCTGACCACTCAGTACGCGCGCGCTTTCGCGGCTTTC 1063
QY 1021 GCAACCGCGCTCTGCTGCTCTTTCCTGAGGCGCGCTGCTGCTGCTGCTGCTGCTGCT 1080
DB 1064 GCAACCGCGCTCTGCTGCTCTTTCCTGAGGCGCGCTGCTGCTGCTGCTGCTGCTGCT 1123
QY 1081 CCCAGCGCTTTCGACACCTTTCGAGCAACAAAGCGCCAGAGACTGCAAGCGAGAGAGG 1140
DB 1124 CCCAGCGCTTTCGACACCTTTCGAGCAACAAAGCGCCAGAGACTGCAAGCGAGAGAGG 1183
QY 1141 GGCTACCTTTCATCTGCGGCAACCACTGCAACAAGAGCGCGCTTCCAGACTTAA 1200
DB 1184 GGCTACCTTTCATCTGCGGCAACCACTGCAACAAGAGCGCGCTTCCAGACTTAA 1243
QY 1201 TGTATACCACTAAGCTGTGAGGGGAGCCCAATCTGACTCTTCCCGGCTTGGAGCAT 1260
DB 1244 TGTATACCACTAAGCTGTGAGGGGAGCCCAATCTGACTCTTCCCGGCTTGGAGCAT 1303

QY 1261 CGCAGCGCGGAGACAGTGGCCGCCAGGCTGGGCCAGAGAGCTCCAGAGGGGCACTG 1320
Db 1304 CGCAGCGCGGAGAGCAGTGGCCGCCAGGCTGGGCCAGAGAGCTCCAGAGGGGCACTG 1363
QY 1321 AGCGCTGCTGGCGCGAGGCTCGACATCCGACAGCACAGGAAAGTCTCTGGGCGCA 1380
Db 1364 AGCGCTGCTGGCGCGAGGCTCGACATCCGACAGCACAGGAAAGTCTCTGGGCGCA 1423
QY 1381 TCTGTAATAAACCTTTTCTTTTCTTTTAAAAA 1438
Db 1424 TCTGTAATAAACCTTTTCTTTTCTTTTAAAAA 1481

RESULT 4
ABL90535
ID ABL90535 standard; cDNA; 1474 BP.
XX ABL90535;
AC
XX
XX 24-MAY-2002 (first entry)
DT Human polynucleotide SEQ ID NO 1097.
DE
XX
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KM antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KM vulnarant; anticonvulsant; antibacterial; antifungal; antiparasitic;
KM cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KM neurological disease; infection; human; secreted protein; gene; ss.
XX
OS Homo sapiens.
XX
XX WO200190304-A2.
XX
XX 29-NOV-2001.
PD
XX
XX 18-MAY-2001; 2001WO-US016450.
PF
XX 19-MAY-2000; 2000US-0205515P.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Birse CE, Rosen CA;
PI
XX WPI; 2002-122018/16.
DR P-PSDB; ABB90126.
DR
XX
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders.
PS
XX Claim 4; SEQ ID NO 1097; 2081bp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing
CC / (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1474 BP; 292 A; 480 C; 404 G; 296 T; 0 U; 2 Other;

Query Match 89.2%; Score 1285; DB 6; Length 1474;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1435; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAGTAACGGCTACAGACAGTGAAGAAATAGTTTCGCTGCGCGGCTAGAAAACTGTGCG 60
Db 27 AAAGTAACGGCTACAGACAGTGAAGAAATAGTTTCGCTGCGCGGCTAGAAAACTGTGCG 86
QY 61 GTACCAACCCAGAGCGGTGAG 120
Db 87 GTACCAACCCAGAGCGGTGAG 146
QY 121 GGAAGTACAGCTTACAG 180
Db 147 GGAAGTACAGCTTACAG 206
QY 181 GCGCTCTCCCGGCTCCAG 240
Db 207 GCGCTCTCCCGGCTCCAG 266
QY 241 ACTGCGCGCGGCTGACAG 300
Db 267 ACTGCGCGCGGCTGACAG 326
QY 301 CGGATATCGCGAG 360
Db 327 CGGATATCGCGAG 386
QY 361 GCAGCAAGCTTCTCTGCTCATTTGCGGAGATCCGTGCGCACTCGCGCTGTTGGTTG 420
Db 387 GCAGCAAGCTTCTCTGCTCATTTGCGGAGATCCGTGCGCACTCGCGCTGTTGGTTG 446
QY 421 GTGAGAGTTCTTCTCAGTCTGTTTCATAGGCGAGAAATGTGGCTGTGCACTTCACTGCA 480
Db 447 GTGAGAGTTCTTCTCAGTCTGTTTCATAGGCGAGAAATGTGGCTGTGCACTTCACTGCA 506
QY 481 GAATGTTCTGTTGTTACAGTGAACACCAACATCTCTCAAAAGCTTTCAGCGAGCGCGC 540
Db 507 GAATGTTCTGTTGTTACAGTGAACACCAACATCTCTCAAAAGCTTTCAGCGAGCGCGC 566
QY 541 GTTACAGCCCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db 567 GTTACAGCCCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 626
QY 601 ACCCAGTGCATCAGCTGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 627 ACCCAGTGCATCAGCTGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 686
QY 661 AAAGAGAATTACCGCGGAG 720
Db 687 AAAGAGAATTACCGCGGAG 746
QY 721 CTCTACCTGCGGAG 780
Db 747 CTCTACCTGCGGAG 806
QY 781 CTGCGGAG 840
Db 807 CTGCGGAG 866
QY 841 AACGTGCTGCTTCCACAGCGCGCGCGCGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Db 867 AACGTGCTGCTTCCACAGCGCGCGCGCGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAG 926
QY 901 GCGTTCGCGCTTCTGCGGAG 960
Db 927 GCGTTCGCGCTTCTGCGGAG 986
QY 961 CTGCGCTAGAGCTCTCCGCGCTACAGCTACAGTACAGAGAGAGAGAGAGAGAGAGAGAG 1020
Db 987 CTGCGCTAGAGCTCTCCGCGCTACAGCTACAGTACAGAGAGAGAGAGAGAGAGAGAGAG 1046
QY 1021 GCAACCGCGCTCTGCTGCTCTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080

Db	1047	GCACCGGCGTCTGTGCTCTTCTCTCGAGGGCCGTGTAGTCTCCAGTATGTTCCG	1106
QY	1081	CCCAGCGCTCTTGCACACCTTCTTGACCAAGCGCCAAAGACTGCAGCCAGAGAGAGG	1140
Db	1107	CCCAGCGCTCTTGCACACCTTCTTGACCAAGCGCCAAAGACTGCAGCCAGAGAGAGG	1166
QY	1141	GGCTCACCTCTTATCTCTCGCGCACCCACTGCACAAGCAGGCGCTCTCCCACTTAAAA	1200
Db	1167	GGCTCACCTCTTATCTCTCGCGCACCCACTGCACAAGCAGGCGCTCTCCCACTTAAAA	1226
QY	1201	TGTATCACCACTAACTGTGAGGGGGACCCCAATCTGGACTCTTCCCGCTTGGACAT	1260
Db	1227	TGTATCACCACTAACTGTGAGGGGGACCCCAATCTGGACTCTTCCCGCTTGGACAT	1286
QY	1261	CGCAGGCGGGGAAGCAGTGCCTCCGACAGGCTGGGCCAGAGAGTCCAGGAAGGCACTG	1320
Db	1287	CGCAGGCGGGGAAGCAGTGCCTCCGACAGGCTGGGCCAGAGAGTCCAGGAAGGCACTG	1346
QY	1321	AGCGTGCTGGCGGAGGCTCGGACATCCGACAGCACACGAGGAAAGTCTCTGGGCGGA	1380
Db	1347	AGCGTGCTGGCGGAGGCTCGGACATCCGACAGCACACGAGGAAAGTCTCTGGGCGGA	1406
QY	1381	TCTGTAATTAACCTTTTTTCTTTTGTTTTAAAAAATTTTTTTTTTTTTTTTTTTTTTTT	1438
Db	1407	TCTGTAATTAACCTTTTTTCTTTTGTTTTAAAAAATTTTTTTTTTTTTTTTTTTTTTTT	1464

RESULT 5
AAC95564

ID AAC95564 standard; cDNA; 1460 BP.

AC AAC95564

DT 21-FEB-2001 (first entry)

DE Human secreted protein gene 44 SEQ ID NO:54.

KW Human; secreted protein; cytostatic; immunosuppressive; nocotropic;
KW neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antitumor; anticonvulsant; antibacterial;
KW antifungal; antiparasitic; cardiant; cancer; immune disease; allergy;
KW cardiovascular disorder; wound healing; infection; neurological disease;
KW ss.

OS Homo sapiens.

PN WO200061596-A1.

PD 19-OCT-2000

PF 06-APR-2000; 2000WO-US008983.

PR 09-APR-1999; 99US-0128703P.

XX

PA (ROSE/) ROSEN C A.

PI Ruben SM, Komatsoulis G;

DR WPI; 2000-611865/58.

PT the prevention, t

XX	6
----	---

XX

CC human secreted pr

CC AAB52103 represent

```

CC amino acid sequences with which they share homology. The genes and
CC proteins have activities dependent on the tissues and cells in which they
CC are expressed. Examples of their activities include cytostatic;
CC immunosuppressive; nootropic; neuroprotective; antiviral; antiallergic;
CC hepatotropic; antidiabetic; antiinflammatory; antitumor; vlnary;
CC anticonvulsant; antibacterial; antifungal; antiparasitic; and cardiac.
CC The secreted proteins, polypeptides, antagonists and agonists may be
CC useful in treating, preventing and/or diagnosing diseases and disorders
CC such as cancer, particularly breast and ovarian cancer, and other cancers
CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,
CC liver, lung, or urogenital. Immune disorders such as Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases such as viral,
CC bacterial, fungal and parasitic infections may also be treated using the
CC proteins and polynucleotides of the invention. Sequences AAC95512-
CC AAC95520 and AAB52011 are used in the isolation and characterisation of
CC the proteins and polynucleotides of the invention
XX
XX Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;
SQ
Query Match          69.1%; Score 996; DB 3; Length 1460;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1046; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      1 AAAAGTAACGGCTACAGACAAGTGAAGAATAGTTTCGCTCGCCGGCTAGAAAACTGTGCG 60
DB      22 AAAGTAACGGCTACAGACAAGTGAAGAATAGTTTCGCTCGCCGGCTAGAAAACTGTGCG 81
QY      61 GTACCAACCAGAGCGCTTGAGAGCAGCCACCTCCACGCTTCTTAACGAGAGGTGCA 120
DB      82 GTACCAACCAGAGCGCTTGAGAGCAGCCACCTCCACGCTTCTTAACGAGAGGTGCA 141
QY      121 GGACTCAGACTTCAACGAGCCCACTCGTCCACGCTTGTACGAAAGAGAGCCCAAGGAC 180
DB      142 GGACTCAGACTTCAACGAGCCCACTCGTCCACGCTTGTACGAAAGAGAGCTCAAGAAC 201
QY      181 GCGCTCTCCCGCGCTCCAGGAGCCCAAGCTTGTGCTGCTGCCTGCCCTGCGTGACAGC 240
DB      202 GCGCTCTCCCGCGCTCCAGGAGCCCAAGCTTGTGCTGCTGCCTGCCCTGCGTGACAGC 261
QY      241 ACTCGGCCGGCGGTGACGACATGACCCCTGTGGAACGGCGTACTGCTTTTACCACAGCCC 300
DB      262 ACTCGGCCGGCGGTGACGACATGACCCCTGTGGAACGGCGTACTGCTTTTACCACAGCCC 321
QY      301 CGGCATGCCGAGGCTTCAGCGTTCACCTGCTCATCGTTATTCTAGTGTTTGGCTCTA 360
DB      322 CGGCATGCCGAGGCTTCAGCGTTCACCTGCTCATCGTTATTCTAGTGTTTGGCTCTA 381
QY      361 GCAGCAAGCTTCCCTGCTCATCTTCCGGGGGATCCGTGGCCACTCGCGCTGTTTGGTTG 420
DB      382 GCAGCAAGCTTCCCTGCTCATCTTCCGGGGGATCCGTGGCCACTCGCGCTGTTTGGTTG 441
QY      421 GTGAGATTCTTCTCAGTCTGTGTTCAATAGGCGCAGAAATTGTGGCTGTGCACTTAGTGCA 480
DB      442 GTGAGATTCTTCTCAGTCTGTGTTCAATAGGCGCAGAAATTGTGGCTGTGCACTTAGTGCA 501
QY      481 GAATGTTGTTGGTGAAGTGAACCAACAACATCTTCAAAAAGCTTTCAGCGCAGCGCGC 540
DB      502 GAATGTTGTTGGTGAAGTGAACCAACAACATCTTCAAAAAGCTTTCAGCGCAGCGCGC 561
QY      541 GTTACAGCCCCGTGCGGTCTGCTCGTGGGCCCTGAGAGGCATTATAATTACACTCACAGGG 600
DB      562 GTTACAGCCCCGTGCGGTCTGCTCGTGGGCCCTGAGAGGCATTATAATTACACTCACAGGG 621
QY      601 ACCCCAGTGCATCAGCTGAACGAGACCATTGACTACAAAGAGAGTTTCACTGCGCTG 660
DB      622 ACCCCAGTGCATCAGCTGAACGAGACCATTGACTACAAAGAGAGTTTCACTGCGCTG 681
QY      661 AAAGAGATTACGCCGCGAGTACGCGAAGCAGCACTGGAGAGGGGCTGCCGAGCCAGTG 720

```


Db 682 AAGAGATTACGCCGCGAGTACGCGAACGCACTGGAGAGGGGCTGCCGACCCAGTg 741
Qy 721 CTCTACCTGGCGGAGAGTTACACCGAGTAGCCCTTGCGGCTGTACCAACAGTACgAC 780
Db 742 CTCTACCTGGCGGAGAGTTACACCGAGTAGCCCTTGCGGCTGTACCAACAGTACgAC 801
Qy 781 CTGGCGGGACACTACGCTCTGGCGCAAGCTATGGGTGGCTTCTGTGCTCTCTCTCC 840
Db 802 CTGGCGGGACACTACGCTCTGGCGCAAGCTATGGGTGGCTTCTGTGCTCTCTCTCC 861
Qy 841 AACGTGTGCTCTCTCCAGCGCGCGCGCTCTACGAGAGGCGCTGCACTGTGACCAACCGGA 900
Db 862 AACGTGTGCTCTCTCCAGCGCGCGCGCTCTACGAGAGGCGCTGCACTGTGACCAACCGGA 921
Qy 901 GCCTTGGCGCTCTTGGGGGCTCTTGGCTTGGCTTCCATCTTACGCGTGGCGCTTGGCGG 960
Db 922 GCCTTGGCGCTCTTGGGGGCTCTTGGCTTGGCTTCCATCTTACGCGTGGCGCTTGGCGG 981
Qy 961 CTCCGCTAGGCTCTCTCCGCGCTACCACTCAGTACGCGCGCGCTTCTGGGTCAAGCTG 1020
Db 982 CTCCGCTAGGCTCTCTCCGCGCTACCACTCAGTACGCGCGCGCTTCTGGGTCAAGCTG 1041
Qy 1021 GCAACCGCGCTCTCTGTCCTCTCTCTC 1047
Db 1042 GCAACCGCGCTCTCTGTCCTCTCTCTC 1068

RESULT 6
ABZ67235

ID ABZ67235 standard; cDNA; 1460 BP.

AC ABZ67235;

DT 26-MAR-2003 (first entry)

DE Human secreted protein encoding cDNA SEQ ID NO 355.

XX Human; secreted protein; nootropic; neuroprotective; cyostatic;
KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW vulnary; antibacterial; antiparkinsonian; antisticking; antianaemic;
KW antidiabetic; cancer; antirheumatic; hepatotropic; cerebroprotective;
KW antiinflammatory; antiallergic; antidiabetic; antidiabetic; anticonvulsant;
KW antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine;
KW cardiovascular disorder; neurological disease; nephrotropic;
KW gene therapy; gene; chromosome 9p21; ds.

OS Homo sapiens.

PN WO200277186-A2.

PD 03-OCT-2002.

PF 26-MAR-2002; 2002WO-US009188.

PR 27-MAR-2001; 2001US-0278650F.

PR 12-SEP-2001; 2001US-00950082.

PR 12-SEP-2001; 2001US-00950083.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR WPI; 2003-040583/03.

DR P-PSDB; ABP99814.

XX New human secreted proteins encoded by genes contained in cDNA clones

PT (e.g. HGCA19), useful for preventing, treating or diagnosing e.g. AIDS,

PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or

PT West Nile fever.

PS Claim 7; Page 1353; 2423pp; English.

CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections
XX Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;
SQ
Query Match 69.1%; Score 996; DB 7; Length 1460;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1046; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 AAGTAAACGCTACAGACAGTGAATAATGTTGCTGCGCGCTAGAAAACTGTGCG 60
Db 22 AAGTAAACGCTACAGACAGTGAATAATGTTGCTGCGCGCTAGAAAACTGTGCG 81
Qy 61 GTACCAACCCAGAGCGTTGAGAGCAGCGCCACCTCCAGCTTCTTAAACGAGAGTGCA 120
Db 82 GTACCAACCCAGAGCGTTGAGAGCAGCGCCACCTCCAGCTTCTTAAACGAGAGTGCA 141
Qy 121 GGAAGTCAAGCTTCAACAGCGCCACTCGGTCCAGCGCTTGTATGCAAGAGAGAGGAC 180
Db 142 GGAAGTCAAGCTTCAACAGCGCCACTCGGTCCAGCGCTTGTATGCAAGAGAGGAC 201
Qy 181 GCGCTCTCCGCGGTCCAGAGCAGCGCCAGCTTGTGCTGCTGCGCGCTGCGAGC 240
Db 202 GCGCTCTCCGCGGTCCAGAGCAGCGCCAGCTTGTGCTGCTGCGCGCTGCGAGC 261
Qy 241 ACTCGCGCGGTGAGAGTACAGTACCGCTGCGAGAGCGCGTACTGCTTTTACCCAGGCC 300
Db 262 ACTCGCGCGGTGAGAGTACAGTACCGCTGCGAGAGCGCGTACTGCTTTTACCCAGGCC 321
Qy 301 CCGCATGCCGAGGCTTCAAGCGTTCAGTCTCATGCTATTTAGTGTGCTCTA 360
Db 322 CCGCATGCCGAGGCTTCAAGCGTTCAGTCTCATGCTATTTAGTGTGCTCTA 381
Qy 361 GAGCAAGCTTCTGCTCATCTTGGCGGGATCCGTGCCACTCGCGCTGTTGGTTG 420
Db 382 GAGCAAGCTTCTGCTCATCTTGGCGGGATCCGTGCCACTCGCGCTGTTGGTTG 441
Qy 421 GTGAGAGTCTTCTCAAGTCTGTTATAGCGCGAGAAATGTGCTGCACTTCACTGCA 480
Db 442 GTGAGAGTCTTCTCAAGTCTGTTATAGCGCGAGAAATGTGCTGCACTTCACTGCA 501
Qy 481 GAATGTTGTTGGGTACAGTGAACACCAACATCTTAAAGCTTCAAGCGAGCGCGC 540
Db 502 GAATGTTGTTGGGTACAGTGAACACCAACATCTTAAAGCTTCAAGCGAGCGCGC 561
Qy 541 GTTACAGCCCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db 562 GTTACAGCCCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 621
Qy 601 ACCCAGTGCATCAGCTGAACAGAGACATGACTACAGAGAGTTCACCTGGCGTCTG 660
Db 622 ACCCAGTGCATCAGCTGAACAGAGACATGACTACAGAGAGTTCACCTGGCGTCTG 681
Qy 661 AAGAGAAATTACGCCGCGAGTACGCGAACGCACTGGAAGAGGGGCTGCCGACCACTG 720
Db 682 AAGAGAAATTACGCCGCGAGTACGCGAACGCACTGGAAGAGGGGCTGCCGACCACTG 741
Qy 721 CTCTACCTGGCGGAGAGTTACACCGAGTAGCCCTTGCGGCTGTACCAACAGTACgAC 780
Db 742 CTCTACCTGGCGGAGAGTTACACCGAGTAGCCCTTGCGGCTGTACCAACAGTACgAC 801

The invention relates to novel human genes (ABZ66891-ABZ68209) and the

QY 781 CTGGCGGACACTACGCTTCGGCCACGCTATAGGATGGCGTTCTGCTTCTGAGCTCCTCC 840
DB 802 CTGGCGGAGACACTACGCTTCGGCCACGCTATAGGATGGCGTTCTGCTTCTGAGCTCCTCC 861
QY 841 AACGTGCTGCTCTCCACGCGCGCCGCTCTACGAGGCGCTGACCTGTGACCCAGCA 900
DB 862 AACGTGCTGCTCTCCACGCGCGCCGCTCTACGAGGCGCTGACCTGTGACCCAGCA 921
QY 901 GCCTTCGCGCTCTTCGGGGGCTTCGCGCTTGGCTTCATCTTAGCGGTGCGCTTGGCCG 960
DB 922 GCCTTCGCGCTCTTCGGGGGCTTCGCGCTTGGCTTCATCTTAGCGGTGCGCTTGGCCG 981
QY 961 CTCGCGCTAGGCTCTCCGCGCTCACCACTACGAGCGCGCGCTTCTGGGTCACGCTG 1020
DB 982 CTCGCGCTAGGCTCTCCGCGCTCACCACTACGAGCGCGCGCTTCTGGGTCACGCTG 1041
QY 1021 GCAACCGGCGTCTGTGCTCTTCTC 1047
DB 1042 GCAACCGGCGTCTGTGCTCTTCTC 1068

RESULT 7

ABZ73640
ID ABZ73640 standard; cDNA; 1460 BP.

AC ABZ73640;

DT 12-MAY-2003 (first entry)

DE Secreted protein-encoding gene 360 cDNA clone HUFCL31, SEQ ID NO:370.

XX Human; secreted protein; cancer; tumour; hyperproliferative disorder;
XX autoimmune disorder; inflammation; angiogenic diseases; AIDS;
XX acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
XX drug screening; chromosome identification; chromosome mapping;
XX cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
XX antianaemic; vulnery; chromosome 9p21; gene; ss.

OS Homo sapiens.

PN WO200277013-A2.

PD 03-OCT-2002.

PF 26-MAR-2002; 2002WO-US009370.

PR 27-MAR-2001; 2001US-0278650P.

PR 12-SEP-2001; 2001US-00950082.

PR 12-SEP-2001; 2001US-00950083.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR WPI; 2003-040578/03.

DR P-PSDB; ABR01306.

PS Claim 21; Page 1345; 2474pp; English.

XX ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
XX protein genes, and ABP00947-ABP01363 represent the proteins they encode.
XX ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
XX invention also encompasses antibodies specific for the secreted proteins,
XX the use of the secreted proteins in drug screening and recombinant
XX vectors and host cells comprising a nucleic acid of the invention. The
XX associated proteins are thought to be involved in biological activities
XX associated with cellular signalling, cellular differentiation, cell
XX migration, prohormone activation and neurotransmitter activity. The
XX secreted proteins, nucleic acids encoding them, antibodies or antibody

CC fragments specific for the secreted proteins, and modulators of protein
CC activity are useful for diagnosing or treating cancers or other
CC hyperproliferative disorders. Additionally, the secreted proteins and
CC their nucleic acids may also be used in the treatment of autoimmune
CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
CC wound healing. Nucleic acids of the invention may be used for chromosome
CC identification, chromosome mapping, in gene therapy, for identifying
CC individuals from minute biological samples, as hybridisation probes, and
CC as molecular weight markers. The present sequence represents a human
CC secreted protein-encoding cDNA clone of the invention
XX

SQ Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;

Query Match 69.1%; Score 996; DB 7; Length 1460;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1046; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAGTAACGGGTACAGACAGTGAAGAATAGTTGCTCGCGGCTAGAAAAAAGCTGTG 60

DB 22 AAAGTAACGGGTACAGACAGTGAAGAATAGTTGCTCGCGGCTAGAAAAAAGCTGTG 81

QY 61 GTACCAACCCAGAGCGTTGAGAGCAGCCCACTCCAGCTTCTTAACGAGAGGTGCA 120

DB 82 GTACCAACCCAGAGCGTTGAGAGCAGCCCACTCCAGCTTCTTAACGAGAGGTGCA 141

QY 121 GGACTCAGACTTACCAAGCCCACTGGTCCAGCTTGTACGCAAGAGAGCGCAAGGAC 180

DB 142 GGACTCAGACTTACCAAGCCCACTGGTCCAGCTTGTACGCAAGAGAGCGCAAGGAC 201

QY 181 GGGCTTCCGCGCTCCAGGAGAGCCCACTGGTCCAGCTTGTACGCAAGAGAGCGCAAGGAC 240

DB 202 GGGCTTCCGCGCTCCAGGAGAGCCCACTGGTCCAGCTTGTACGCAAGAGAGCGCAAGGAC 261

QY 241 ACTCGCGCGCGTGCAGATGACCCCTGTGGAACGCGCTACTGCTTTTACCACCGCC 300

DB 262 ACTCGCGCGCGTGCAGATGACCCCTGTGGAACGCGCTACTGCTTTTACCACCGCC 321

QY 301 CGGATGCGCGAGGCTTACGCGTTCACCTGCTCATGTTATTCTAGTGTGCTCTA 360

DB 322 CGGATGCGCGAGGCTTACGCGTTCACCTGCTCATGTTATTCTAGTGTGCTCTA 381

QY 361 GCAAGCAAGCTTCTCTCATCTTGGCGGAGATCCGTGGCCACTCGCGCTGTTGGTTG 420

DB 382 GCAAGCAAGCTTCTCTCATCTTGGCGGAGATCCGTGGCCACTCGCGCTGTTGGTTG 441

QY 421 GTGAGAGTTCTTCTCAGTCTGTTCATAGGCGCAAGAAATGCTGTGCACTTCAGTGCA 480

DB 442 GTGAGAGTTCTTCTCAGTCTGTTCATAGGCGCAAGAAATGCTGTGCACTTCAGTGCA 501

QY 481 GAATGTTCTGTGGTACAGTGAACCAACACATCTCTACAAAGCCTTACGCGAGCGGC 540

DB 502 GAATGTTCTGTGGTACAGTGAACCAACACATCTCTACAAAGCCTTACGCGAGCGGC 561

QY 541 GTTACAGCCCGTGTGCTGCTGCTGGGCTGAGGGGATATATATTACTACAGAGG 600

DB 562 GTTACAGCCCGTGTGCTGCTGCTGGGCTGAGGGGATATATATTACTACAGAGG 621

QY 601 ACCCAGTGCATCAGCTGAACGAGACCATTAACAAGAGCAGATTGACCTGGCGTGG 660

DB 622 ACCCAGTGCATCAGCTGAACGAGACCATTAACAAGAGCAGATTGACCTGGCGTGG 681

QY 661 AAAGAGATTAAGCGCGGAGTACGCGCAAGCGCACTGAGAGGGGCTGCGGACCCAGTG 720

DB 682 AAAGAGATTAAGCGCGGAGTACGCGCAAGCGCACTGAGAGGGGCTGCGGACCCAGTG 741

QY 721 CTCTACCTGGCGGAGAGTTACACCGAGTACGCTTGGCGCTGTACACAGTACAC 780

DB 742 CTCTACCTGGCGGAGAGTTACACCGAGTACGCTTGGCGCTGTACACAGTACAC 801

QY 781 CTGGCGGACACTACGCTTCGGCCACGCTATAGGATGGCGTTCTGCTTCTGAGCTCCTCC 840

DB 802 CTGGCGGACACTACGCTTCGGCCACGCTATAGGATGGCGTTCTGCTTCTGAGCTCCTCC 861

QY 841 AACGTGCTGCTCTCCACGCGCGCCGCTCTACGAGGCGCTGCACTGTGACCAACCGGA 900
DB 862 AACGTGCTGCTCTCCACGCGCGCCGCTCTACGAGGCGCTGCACTGTGACCAACCGGA 921
QY 901 GCCTTCGCGCTCTTCGGGGTCTTCGGCTTGGCTTCCATCTCTAGCGTGCCTCTGCCCC 960
DB 922 GCCTTCGCGCTCTTCGGGGTCTTCGGCTTGGCTTCCATCTCTAGCGTGCCTCTGCCCC 981
QY 961 CTCGCGCTAGGCTCTCCGCGCTCAACCACTAGTACGCGCGCGCTTCTGGGTACGCTG 1020
DB 982 CTCGCGCTAGGCTCTCCGCGCTCAACCACTAGTACGCGCGCGCTTCTGGGTACGCTG 1041
QY 1021 GCAACCGCGCTCTGCTCTTCTC 1047
DB 1042 GCAACCGCGCTCTGCTCTTCTC 1068
RESULT 8
ID ADC20289 standard; DNA; 1460 BP.
AC ADC20289;
DT 18-DEC-2003 (first entry)
DE Human secreted protein coding sequence #228.
XX
KW gene therapy; human; secreted protein; haemopoietic disorder;
KW haematological disorder; anaemia; haemophilia; inflammatory disorder;
KW inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
KW leukaemia; wound healing; epithelial cell proliferation disorder;
KW immune disorder; autoimmune disorder; asthmatic disorder;
KW cardiovascular disorder; atherosclerosis; myocarditis;
KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;
KW gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200292787-A2.
PD 21-NOV-2002.
XX
PF 26-MAR-2002; 2002WC-US009257.
XX
PR 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-129287/12.
XX
PT New human secreted proteins and nucleic acid molecules, useful for
PT preparing a diagnostic or pharmaceutical composition for diagnosing,
PT preventing or treating hematopoietic or hematologic disorders, e.g.
PT anemia or hemophilia.
XX
PS Claim 1; SEQ ID NO 238; 1512pp; English.
XX
CC The invention comprises the amino acid and coding sequences of human
CC secreted proteins. The DNA and protein sequences of the invention are
CC useful for detecting, preventing, diagnosing, prognosticating, treating
CC or ameliorating: haematopoietic or haematological disorders (e.g. anaemia
CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease
CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
CC wound healing and disorders of epithelial cell proliferation; immune
CC disorders (e.g. autoimmune disorders and asthmatic disorders);
CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);
CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
CC and gastrointestinal disorders (e.g. duodenal ulcers and

CC gastroenteritis). The present DNA sequence encodes a human secreted
CC protein of the invention.
XX
SQ Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;
Query Match 69.1%; Score 996; DB 9; Length 1460;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1046; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAGTAAACGGCTACAGACAGTGAAGAATGTTTGGCTGCGCGGCTAGAAAACTCTGTG 60
DB 22 AAGTAAACGGCTACAGACAGTGAAGAATGTTTGGCTGCGCGGCTAGAAAACTCTGTG 81
QY 61 GTACCAACCCGAGAGCGTTGAGAGCAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 120
DB 82 GTACCAACCCGAGAGCGTTGAGAGCAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 141
QY 121 GGACTCAGACTTCAACCAAGCCCACTCGTCCCAAGCTTGTACGCAAGAGAGCAGCA 180
DB 142 GGACTCAGACTTCAACCAAGCCCACTCGTCCCAAGCTTGTACGCAAGAGAGCAGCA 201
QY 181 GCGCTCTCCCGCTCCAGGAGCCCAAGCTTGTGCTGCTGCGCGCTGCGTGCAGC 240
DB 202 GCGCTCTCCCGCTCCAGGAGCCCAAGCTTGTGCTGCTGCGCGCTGCGTGCAGC 261
QY 241 ACTCGCGCGCGCTGAGCAGATGACCTGTGGAACGCGCTACTGCTTTTAAACCCAGCC 300
DB 262 ACTCGCGCGCGCTGAGCAGATGACCTGTGGAACGCGCTACTGCTTTTAAACCCAGCC 321
QY 301 CGGATCGCGCAGGCTTACGCTTACGCTTACGCTTATTTAGTGTGCTCTA 360
DB 322 CGGATCGCGCAGGCTTACGCTTACGCTTACGCTTATTTAGTGTGCTCTA 381
QY 361 GCAGCAAGCTTCTGCTCATCTTGCCGGGATCCGTGCGCACTCGCGCTGTTGTTG 420
DB 382 GCAGCAAGCTTCTGCTCATCTTGCCGGGATCCGTGCGCACTCGCGCTGTTGTTG 441
QY 421 GTGAGAGTCTTCTCAGTCTGTTCATAGGCGCAGAAATTTGGCTGTGCACTTCA 480
DB 442 GTGAGAGTCTTCTCAGTCTGTTCATAGGCGCAGAAATTTGGCTGTGCACTTCA 501
QY 481 GAATGTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
DB 502 GAATGTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 561
QY 541 GTTACAGCCCGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 600
DB 562 GTTACAGCCCGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 621
QY 601 ACCCAGTGCATCAGCTGAACGAGACCATTTACTACAACGACAGTTCACTGGCGTCTG 660
DB 622 ACCCAGTGCATCAGCTGAACGAGACCATTTACTACAACGAGTTCACTGGCGTCTG 681
QY 661 AAAGAGATTACGCGCGGAGTACGCGAAACGCACTGAGAGGGGCTGCCGACCCAGTG 720
DB 682 AAAGAGATTACGCGCGGAGTACGCGAAACGCACTGAGAGGGGCTGCCGACCCAGTG 741
QY 721 CTCTACCTGGCGGAGAGTTACACAGGAGTACCTTGGCGCTGTACCAACAGTACAC 780
DB 742 CTCTACCTGGCGGAGAGTTACACAGGAGTACCTTGGCGCTGTACCAACAGTACAC 801
QY 781 CTGCGGAGACACTACGCTCGGCAAGCTATAGGTGGCTTCTGCTCTCTCTCTCTCT 840
DB 802 CTGCGGAGACACTACGCTCGGCAAGCTATAGGTGGCTTCTGCTCTCTCTCTCTCTCT 861
QY 841 AACGTGCTCTTCCACGCGCGCGCTCTACGAGGCGCTGCACTGTGCAACCGGA 900
DB 862 AACGTGCTCTTCCACGCGCGCGCTCTACGAGGCGCTGCACTGTGCAACCGGA 921
QY 901 GCCTTCGCGCTCTTCGGGGTCTTCGGCTTGGCTTCCATCTCTAGCGTGCCTCTGCCCC 960
DB 922 GCCTTCGCGCTCTTCGGGGTCTTCGGCTTGGCTTCCATCTCTAGCGTGCCTCTGCCCC 981

QY 961 CTCGGCCTAGGCTCCTCCGGCTCACCACTCAGTACGCGCGCCCTTCTGGGTACAGCTG 1020
 DB 982 CTCGGCCTAGGCTCCTCCGGCTCACCACTCAGTACGCGCGCCCTTCTGGGTACAGCTG 1041
 QY 1021 GCAACCGGCGCTCCTGTGCTCTTCTCTC 1047
 DB 1042 GCAACCGGCGCTCCTGTGCTCTTCTCTC 1068

RESULT 9

AAA94624

ID AAA94624 standard; DNA; 498 BP.

AC AAA94624;

DT 11-JAN-2001 (first entry)

DE Human CASB618 EST.

KW Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian;
 KW colon; autoimmune disease; HLA_A0201; expressed sequence tag; EST; ss.

OS Homo sapiens.

PN WO200053748-A2.

PD 14-SEP-2000.

PF 09-MAR-2000; 2000WO-EP002048.

PR 11-MAR-1999; 99GB-00005607.

PR 01-SEP-1999; 99GB-00020590.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Bruck CEM, Cassart J, Coche T, Vinals Y De Bassolac;

PI WPI; 2000-572268/53.

PT New human CASB618 polypeptide, useful as a vaccine for prophylactic and
 PT therapeutic treatment of cancers, particularly ovarian or colon cancer,
 PT autoimmune diseases and related conditions.

PS Claim 32; Page 62; 76pp; English.

CC The present sequence is an expressed sequence tag (EST) for human CASB618
 CC protein. The gene for human CASB618 is thought to be located on
 CC chromosome 15. CASB618 protein and epitopes of CASB618 protein (see
 CC AAB26327 to AAB26399) are useful in diagnosing the occurrence of tumour
 CC cells and in vaccines for prophylactic and therapeutic treatment of
 CC cancers, particularly ovarian or colon cancer, autoimmune diseases and
 CC related conditions

SQ Sequence 498 BP; 97 A; 166 C; 137 G; 98 T; 0 U; 0 Other;

Query Match 29.5%; Score 425; DB 3; Length 498;

Best Local Similarity 100.0%; Pred. No. 1.9e-153; Indels 0; Gaps 0;

Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 998 GCGCCGCTTCTGGGTACCGCTGGCAACCGCGCTGCTGTGCTCTTCTCTCGAGGGCGCG 1057
 DB 62 GCGCCGCTTCTGGGTACCGCTGGCAACCGCGCTGCTGTGCTCTTCTCTCGAGGGCGCG 121

QY 1058 TGGTGAGTCTTCAGTATGTTGGCCCGAGGCTCTTCGCACCCCTTTCGACCAAGCGCCA 1117
 DB 122 TGGTGAGTCTTCAGTATGTTGGCCCGAGGCTCTTCGCACCCCTTTCGACCAAGCGCCA 181

QY 1118 AGGACTGCAGGAGAGAGAGAGGGGGCTCACTCTTATCTCGGCAACCCACTGCACAAGC 1177
 DB 182 AGGACTGCAGGAGAGAGAGAGGGGGCTCACTCTTATCTCGGCAACCCACTGCACAAGC 241

QY 1178 AGGCGCTCTCCAGACTTAATATGATACCACTAACCTGTGAGGGGAGCCCAATCTTG 1237
 |||||

DB 242 AGCGCGCTCTCCAGACTTAAATGATATCACCACCTAACCCTGTGAGGGGAGCCCAATCTGG 301
 QY 1238 ACTCCTTCCCGCCTTGGGACATCGCAGGCGCGGAGAGCAGTGCCCCCGCAGGCTGGGCCA 1297
 DB 302 ACTCCTTCCCGCCTTGGGACATCGCAGGCGCGGAGAGCAGTGCCCCCGCAGGCTGGGCCA 361
 QY 1298 GGAGAGCTCCAGGAGAGGCACTGAGCGCTGCTGCGCGGAGGCGCTTCGACATCCGACAGGCA 1357
 DB 362 GGAGAGCTCCAGGAGAGGCACTGAGCGCTGCTGCGCGGAGGCGCTTCGACATCCGACAGGCA 421
 QY 1358 CCAGGGAAGTCTCCTGGGGCGATCTGTAATAAACCTTTTCTTTTGTGTTTAAAA 1417
 DB 422 CCAGGGAAGTCTCCTGGGGCGATCTGTAATAAACCTTTTCTTTTGTGTTTAAAA 481
 QY 1418 AAAA 1422
 DB 482 AAAA 486

RESULT 10

ADD19187

ID ADD19187 standard; cDNA; 580 BP.

AC ADD19187;

DT 15-JAN-2004 (first entry)

DE Human cDNA from secreted protein gene 4.

KW human secreted protein; cytostatic; antibacterial; virucide;
 KW neuroprotective; gynaecological; gastrointestinal; cardiac;
 KW cardiovascular; nephrotropic; antiinflammatory; muscular;
 KW respiratory; immunosuppressive; cerebroprotective; vasotropic;
 KW nootropic; antiallergic; cancer; bacterial infection; viral infection;
 KW neural disorder; immune system disorder; blood disorder;
 KW muscular disorder; reproductive disorder; gastrointestinal disorder;
 KW pulmonary disorder; cardiovascular disorder; renal disorder;
 KW inflammatory disorder; proliferative disorder; human; ss; gene.

OS Homo sapiens.

PN WO2003052377-A2.

PD 26-JUN-2003.

PF 06-NOV-2002; 2002WO-US035606.

PR 07-NOV-2001; 2001US-0331046P.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

PI WPI; 2003-533050/50.

PI P-PSDB; ADD19262.

PT New isolated nucleic acids encoding signal transduction pathway component
 PT polypeptides, useful for diagnosing, treating, and/or preventing
 PT disorders, such as cancer, infections, cardiovascular and inflammatory
 PT diseases.

PS Claim 1; SEQ ID NO 14; 554pp; English.

CC The invention relates to an isolated nucleic acid molecule (cDNA)
 CC encoding a human secreted protein, representing one of 85 novel genes.
 CC Also included are recombinant vectors, host cells (expressing the
 CC protein), the secreted proteins (including their fragments, epitopes and
 CC homologues), an isolated antibody that binds specifically to the protein,
 CC diagnosing a pathological condition or susceptibility to a pathological
 CC condition (comprising determining the presence or absence of a mutation
 CC in the nucleic acid and diagnosing a condition based on the presence or
 CC absence of the mutation), diagnosing a pathological condition or
 CC susceptibility to a pathological condition (comprising determining the

CC presence or amount of expression of the protein in a biological sample
CC and diagnosing a condition based on the presence or amount of expression
CC of the protein), preventing, treating or ameliorating a medical condition
CC by administering the nucleic acid or protein to a mammalian subject,
CC identifying a binding partner to the protein, the gene corresponding to
CC the cDNA sequence, and identifying an activity in a biological assay
CC (comprising expressing the nucleic acid in a cell, isolating the
CC supernatant, detecting an activity in a biological assay and identifying
CC the protein in the supernatant having the activity). The nucleic acids
CC and proteins display the following activities: Cytostatic, antibacterial,
CC Virucide, Neuroprotective, Gynaecological, Gastrointestinal-Gen,
CC Cardiant, Cardiovascular-Gen, Nephrotropic, Antiinflammatory, Muscular-
CC Gen, Respiratory-Gen, Immunosuppressive, Cerebroprotective, Vasotropic,
CC Nootropic, Antiallergic. The methods and compositions of the present
CC invention are useful for diagnosing, treating, preventing and/or
CC prognosticating disorders related to the novel polypeptides, such as
CC cancer, bacterial or viral infections, and neural, immune system, blood,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal, inflammatory or proliferative disorders (many examples of these
CC diseases and disorders are given in the specification). The present
CC sequence encodes a novel secreted protein of the invention.

XX
SQ Sequence 580 BP; 115 A; 198 C; 156 G; 111 T; 0 U; 0 Other;

Query Match 28.1%; Score 405; DB 9; Length 580;
Best Local Similarity 100.0%; Pred. No. 8.6e-146;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTAAACGGCTACAGACAGTGAAGAAATAGTTTGGCTGCGCCGGCTAGAAAACTGTGCG 60
Db 13 AAAGTAAACGGCTACAGACAGTGAAGAAATAGTTTGGCTGCGCCGGCTAGAAAACTGTGCG 72
QY 61 GTACCAACCCAGAGCGTTGAGAGAGAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 120
Db 73 GTACCAACCCAGAGCGTTGAGAGAGAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 132
QY 121 GGAAGTAAACGGCTACAGACAGTGAAGAAATAGTTTGGCTGCGCCGGCTAGAAAACTGTGCG 180
Db 133 GGAAGTAAACGGCTACAGACAGTGAAGAAATAGTTTGGCTGCGCCGGCTAGAAAACTGTGCG 192
QY 181 GCGCTCTCCCGGCTCCAGGAGAGCCCACTGCTGCTGCTGCGCCGGCTAGAAAACTGTGCG 240
Db 193 GCGCTCTCCCGGCTCCAGGAGAGCCCACTGCTGCTGCTGCGCCGGCTAGAAAACTGTGCG 252
QY 241 ACTGCGCGCGGTGAGAGAGTGAAGAAATAGTTTGGCTGCGCCGGCTAGAAAACTGTGCG 300
Db 253 ACTGCGCGCGGTGAGAGAGTGAAGAAATAGTTTGGCTGCGCCGGCTAGAAAACTGTGCG 312
QY 301 CGGCATGCGCGGTGAGAGAGTGAAGAAATAGTTTGGCTGCGCCGGCTAGAAAACTGTGCG 360
Db 313 CGGCATGCGCGGTGAGAGAGTGAAGAAATAGTTTGGCTGCGCCGGCTAGAAAACTGTGCG 372
QY 361 GCAGCAAGCTTCTGCTCATCTGCGCGGAGATCCGTGCGCACTCG 405
Db 373 GCAGCAAGCTTCTGCTCATCTGCGCGGAGATCCGTGCGCACTCG 417

RESULT 11

ABZ68115
ID ABZ68115 standard; DNA; 5033 BP.

XX ABZ68115;

XX 26-MAR-2003 (first entry)

DE Human secreted protein encoding genomic DNA SEQ ID NO 1638.

XX Human; secreted protein; nootropic; neuroprotective; cytosstatic;
KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW vulnerary; antibacterial; antiparkinsonian; antislaking; antianemic;
KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
KW antiinflammatory; antiallergic; antidiabetic; antilucer; anticonvulsant;
KW antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;

KW cardiovascular disorder; neurological disease; nephrotropic;
KW gene therapy; gene; ds.

XX Homo sapiens.

XX WO200277186-A2.

PN 03-OCT-2002.

PD 26-MAR-2002; 2002WO-US009188.

PF 27-MAR-2001; 2001US-0278650P.

PR 12-SEP-2001; 2001US-00950082.

PR 12-SEP-2001; 2001US-00950083.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2003-040583/03.

XX New human secreted proteins encoded by genes contained in cDNA clones

XX (e.g. HGCAC19), useful for preventing, treating or diagnosing e.g. AIDS,

XX multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or

XX West Nile fever.

XX Disclosure; Page 2263-2264; 2423pp; English.

CC The invention relates to novel human genes (ABZ66891-ABZ68209) and the
CC encoded secreted proteins (ABP93470-ABP93872) useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections

XX Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;

Query Match 28.1%; Score 405; DB 7; Length 5033;
Best Local Similarity 100.0%; Pred. No. 6e-146;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTAAACGGCTACAGACAGTGAAGAAATAGTTTGGCTGCGCCGGCTAGAAAACTGTGCG 60
Db 13 AAAGTAAACGGCTACAGACAGTGAAGAAATAGTTTGGCTGCGCCGGCTAGAAAACTGTGCG 72
QY 61 GTACCAACCCAGAGCGTTGAGAGAGAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 120
Db 73 GTACCAACCCAGAGCGTTGAGAGAGAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 132
QY 121 GGAAGTAAACGGCTACAGACAGTGAAGAAATAGTTTGGCTGCGCCGGCTAGAAAACTGTGCG 180
Db 133 GGAAGTAAACGGCTACAGACAGTGAAGAAATAGTTTGGCTGCGCCGGCTAGAAAACTGTGCG 192
QY 181 GCGCTCTCCCGGCTCCAGGAGAGCCCACTGCTGCTGCTGCGCCGGCTAGAAAACTGTGCG 240
Db 193 GCGCTCTCCCGGCTCCAGGAGAGCCCACTGCTGCTGCTGCGCCGGCTAGAAAACTGTGCG 252
QY 241 ACTGCGCGCGGTGAGAGAGTGAAGAAATAGTTTGGCTGCGCCGGCTAGAAAACTGTGCG 300
Db 253 ACTGCGCGCGGTGAGAGAGTGAAGAAATAGTTTGGCTGCGCCGGCTAGAAAACTGTGCG 312
QY 301 CGGCATGCGCGGTGAGAGAGTGAAGAAATAGTTTGGCTGCGCCGGCTAGAAAACTGTGCG 360
Db 313 CGGCATGCGCGGTGAGAGAGTGAAGAAATAGTTTGGCTGCGCCGGCTAGAAAACTGTGCG 372

XX Disclosure; SEQ ID NO 959; 1512pp; English.
PS
XX
CC The invention comprises the amino acid and coding sequences of human
CC secreted proteins. The DNA and protein sequences of the invention are
CC useful for detecting, preventing, diagnosing, prognosticating, treating
CC or ameliorating: haematopoietic or haematological disorders (e.g. anaemia
CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease
CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
CC wound healing and disorders of epithelial cell proliferation; immune
CC disorders (e.g. autoimmune disorders and asthmatic disorders);
CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);
CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
CC and gastrointestinal disorders (e.g. duodenal ulcers and
CC gastroenteritis). The present DNA sequence was used in the
CC exemplification of the invention.
XX
SQ Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;

Query Match	28.1%	Score 405;	DB 9;	Length 5033;
Best Local Similarity	100.0%;	Pred. No. 6e-146;		
Matches 405;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	AAAGTAACGGGCTACAGACAGTGAGAAATAGTTTCGCTCGCCGGCTAGAAAACCTCTGTGC	60
Db	13	AAAGTAACGGGCTACAGACAGTGAGAAATAGTTTCGCTCGCCGGCTAGAAAACCTCTGTGC	72
QY	61	GTACCAACCCCCAGAGCGCTTGAGAGAGAGCCCACTCCACGCTTCCTTAACGAGAGGTGCA	120
Db	73	GTACCAACCCCCAGAGCGCTTGAGAGAGAGCCCACTCCACGCTTCCTTAACGAGAGGTGCA	132
QY	121	GGACTCAGACTTCACCAAGCCCACTCGGTCCAGCGCTGTAGCGAAAGACGCCCAAGAC	180
Db	133	GGACTCAGACTTCACCAAGCCCACTCGGTCCAGCGCTGTAGCGAAAGACGCCCAAGAC	192
QY	181	GCGCTCTCCCGCGTCCAGGACGCCCAAGCTTGCTGGCTGCCCGCGCTGCGTGACG	240
Db	193	GCGCTCTCCCGCGTCCAGGACGCCCAAGCTTGCTGGCTGCCCGCGCTGCGTGACG	252
QY	241	ACTCGGCGGCGGTGACAGATGACCCCTGTGAAACGCGTACTGCTTTTTPAACCCCAAGCC	300
Db	253	ACTCGGCGGCGGTGACAGATGACCCCTGTGAAACGCGTACTGCTTTTTPAACCCCAAGCC	312
QY	301	CGGCATGCGCGAGGCTTCAGCGTTCCACTGCTCATCGTTATTTAGTGTGCTGCTA	360
Db	313	CGGCATGCGCGAGGCTTCAGCGTTCCACTGCTCATCGTTATTTAGTGTGCTGCTA	372
QY	361	GCAGCAAGCTTCTGCTCATCTTGCGGGGATCCGTGGCCACTCG	405
Db	373	GCAGCAAGCTTCTGCTCATCTTGCGGGGATCCGTGGCCACTCG	417

RESULT 14	
ADA53627/c	
ID	ADA53627 standard; cDNA; 2684 BP.
XX	
AC	ADA53627;
XX	
DT	20-NOV-2003 (first entry)
XX	
DE	Human coding sequence, SEQ ID 1195.
XX	
KW	Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW	Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW	inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	EP1293569-A2.
XX	
PD	19-MAR-2003.
XX	
PF	21-MAR-2002; 2002EP-00006586.

XX 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-395539/38.
DR P-PSDB; ADA55265.
XX
XX New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 1; SEQ ID NO 1195; 205pp; English.

XX Sequence 2684 BP, 599 A, 785 C, 684 G, 616 T, 0 U, 0 Other;

Query Match	27.2%;	Score 392;	DB 7;	Length 2684;
Best Local Similarity	100.0%;	Pred. NO. 6.5e-11;		
Matches 392;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1027	GGCGTCTGTGCTCTTCTCTTGAGAGGGGCGGTGTGAGTCTCCAGTATGTTCCGCCAGC	1086
Db	2343	GGCGTCTGTGCTCTTCTCTTGAGAGGGGCGGTGTGAGTCTCCAGTATGTTCCGCCAGC	2284
QY	1087	GCTCTTGCACCCCTTCTGTGACCAAGCGCCAAAGACTGCAGCCAGAGAGAGGGGCTCA	1146
Db	2283	GCTCTTGCACCCCTTCTGTGACCAAGCGCCAAAGACTGCAGCCAGAGAGAGGGGCTCA	2224
QY	1147	CTCTTATCCTCGGCGACCACTGCACAAAGCGGCTCTCCAGACTTAAATGTATC	1206
Db	2223	CTCTTATCCTCGGCGACCACTGCACAAAGCGGCTCTCCAGACTTAAATGTATC	2164
QY	1207	ACCACTAACCTGTGAGGGGGACCCAATCTGSACTCCTTCCCGCCTTGGAATCGCAGG	1266
Db	2163	ACCACTAACCTGTGAGGGGGACCCAATCTGSACTCCTTCCCGCCTTGGAATCGCAGG	2104
QY	1267	CCGGGAAGCAGTGCCTCCGCCAGGCTGGGGCCAGAGAGCTCCAGGAAGGGCACTGAGCGCT	1326
Db	2103	CCGGGAAGCAGTGCCTCCGCCAGGCTGGGGCCAGAGAGCTCCAGGAAGGGCACTGAGCGCT	2044
QY	1327	GCTGGCGCGAGGCTCTGGACATCCGACAGGCACCAGGGAAGTCTCCTGGGGCGATCTGTA	1386
Db	2043	GCTGGCGCGAGGCTCTGGACATCCGACAGGCACCAGGGAAGTCTCCTGGGGCGATCTGTA	1984
QY	1387	AATAAACCTTTTTTCTTTTGTTTTTAAAAA 1418	
Db	1983	AATAAACCTTTTTTCTTTTGTTTTTAAAAA 1952	

RESULT 15
AAS66563/c
ID AAS66563 standard; cDNA; 406 BP.
XX
AC AAS66563;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #2367.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss

XX	Homo sapiens.
OS	
XX	
XX	
PN	WO200175067-A2.
XX	
XX	
PD	11-OCT-2001.
XX	
XX	
PF	30-MAR-2001; 2001WO-US008631.
XX	
PR	31-MAR-2000; 2000US-00540217.
PR	23-AUG-2000; 2000US-00649167.
XX	
XX	
PA	(HYSE-) HYSEQ INC.
XX	
XX	
PI	Dymanac RT, Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73.
DR	P-PSDB; ABG02376.
XX	
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity.

Claim 1; SEQ ID NO 2367; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 406 BP; 83 A; 107 C; 138 G; 78 T; 0 U; 0 Other;

Query Match 21.0%; Score 303; DB 5; Length 406;

Best Local Similarity	99.5%	Pred. No. 1.3e-106;							
Matches	403;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;

QY	1	AAAGTAA	CGGCTACAGACAGT	GAGAAATAGTTTGCTCGCCGCGCTAGAAAAA	CTGTG	60				
Db	405	AAAGTAA	CGGCTACAGACAGT	GAGAAATAGTTTGCTCGCCGCGCTAGAAAAA	CTGTG	346				
QY	61	GTA	CCAA	CCCCAGAGCGCTTGA	GAGCA	CCCCACCTTCC	TAA	CGGAGAGTGA	120	
Db	345	GGA	CCAA	CCCCAGAGCGCTTGA	GAGCA	CCCCACCTTCC	TAA	CGGAGAGTGA	286	
QY	121	GGA	CTCAGACTTCA	CCAGCCCACTCGGTCC	CAGCCTTGTACGCA	AAAGAGCGCA	AGGAC	180		
Db	285	GGA	CTCAGACTTCA	CCAGCCCACTCGGTCC	CAGCCTTGTACGCA	AAAGAGCGCA	AGGAC	226		
QY	181	GCG	CTTCCCGCGTCC	AGGCA	GCCCCAGCTTGTCTG	CGCTTGCCCGCTG	CGTGACG	240		
Db	225	GCG	CTTCCCGCGTCC	AGGCA	GCCCCAGCTTGTCTG	CGCTTGCCCGCTG	CGTGACG	166		
QY	241	ACT	CGGCGCGCGTG	CAGCATGAC	CCCTGTGGAA	CGGCGTACTG	CTTTT	TACCCCA	GAGCC	300
Db	165	ACT	CGGCGCGCGTG	CAGCATGAC	CCCTGTGGAA	CGGCGTACTG	CTTTT	TACCCCA	GAGCC	106

QY	301	CGGCATGCCGCAGGCGTTCAGCGTTCACCTGCTCATCGTATTCTAGTGTGTTTGACTCTA	360
Db	105	CGGCATGCCGCAGGCGTTCAGCGTTCACCTGCTCATCGTATTCTAGTGTGTTTGACTCTA	46
QY	361	GCAGCAAGCTTCCTGCTCATCTTGCCTGGGGATCCGTGCCCACTCG	405
Db	45	GCAGCAAGCTTCCTGCTCATCTTGCCTGGGGATCCGTGCCCACTCG	1

Search completed: February 23, 2004, 23:49:19
Job time : 586.643 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 23:23:40 ; Search time 130.797 Seconds
(without alignments)
6113.919 Million cell updates/sec

Title: US-09-936-456-1

Perfect score: 1441

Sequence: 1 aaagtaacggtacagacag.....aaaaaaaaaaaaaaaa 1441

Scoring table: OLIGO_MUC

Gapop_60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: Issued Patents NA: *
2: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/BACKFILES1.seq: *

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	2.5	2255	3 US-08-871-572B-3	Sequence 3, Appli
2	35	2.4	1639	3 US-09-362-473-5	Sequence 5, Appli
3	34	2.4	144	1 US-08-702-344-26	Sequence 26, Appli
4	34	2.4	348	4 US-09-621-976-13740	Sequence 13740, A
5	34	2.4	388	4 US-09-621-976-18573	Sequence 18573, A
6	34	2.4	396	4 US-09-640-173-10	Sequence 10, Appli
7	34	2.4	396	4 US-09-713-550-10	Sequence 10, Appli
8	34	2.4	397	4 US-09-621-976-18571	Sequence 18571, A
9	34	2.4	769	4 US-08-567-882-5	Sequence 5, Appli
10	34	2.4	1191	3 US-09-282-305-13	Sequence 13, Appli
11	34	2.4	1191	4 US-09-883-720-13	Sequence 13, Appli
12	34	2.4	1283	1 US-08-174-467-19	Sequence 19, Appli
13	34	2.4	1283	3 US-08-452-071-19	Sequence 19, Appli
14	34	2.4	1518	4 US-09-614-912-191	Sequence 191, Appli
15	34	2.4	1582	3 US-08-545-196B-10	Sequence 10, Appli
16	34	2.4	1582	3 US-08-545-196B-12	Sequence 12, Appli
17	34	2.4	2043	4 US-09-614-912-171	Sequence 171, Appli
18	34	2.4	2103	4 US-09-489-847-40	Sequence 40, Appli
19	34	2.4	2218	2 US-08-985-090-4	Sequence 4, Appli
20	34	2.4	2218	3 US-09-165-543-31	Sequence 31, Appli
21	34	2.4	2780	4 US-09-489-847-87	Sequence 87, Appli
22	34	2.4	3244	3 US-09-165-543-4	Sequence 4, Appli
23	34	2.4	3437	3 US-08-860-339-17	Sequence 17, Appli
24	34	2.4	3437	4 US-09-573-629-17	Sequence 17, Appli
25	34	2.4	3581	2 US-08-738-349-1	Sequence 1, Appli
26	34	2.4	4494	4 US-09-620-312D-861	Sequence 861, Appli
27	33	2.3	229	4 US-09-621-976-10914	Sequence 10914, A

ALIGNMENTS

28	33	2.3	273	4 US-09-621-976-11078	Sequence 11078, A
29	33	2.3	375	3 US-08-946-026-23	Sequence 23, Appli
30	33	2.3	724	3 US-09-020-956-38	Sequence 38, Appli
31	33	2.3	724	3 US-09-030-607-38	Sequence 38, Appli
32	33	2.3	724	4 US-09-439-313-38	Sequence 38, Appli
33	33	2.3	724	4 US-09-352-616A-38	Sequence 38, Appli
34	33	2.3	724	4 US-09-232-149A-38	Sequence 38, Appli
35	33	2.3	724	4 US-09-159-812-38	Sequence 38, Appli
36	33	2.3	724	4 US-09-636-215-38	Sequence 38, Appli
37	33	2.3	724	4 US-09-685-166A-38	Sequence 38, Appli
38	33	2.3	724	4 US-09-115-453-38	Sequence 38, Appli
39	33	2.3	724	4 US-09-688-489-38	Sequence 38, Appli
40	33	2.3	911	2 US-08-924-759-9	Sequence 9, Appli
41	33	2.3	911	3 US-09-248-335-9	Sequence 9, Appli
42	33	2.3	1052	4 US-09-489-847-23	Sequence 23, Appli
43	33	2.3	1223	3 US-09-154-874-4	Sequence 4, Appli
44	33	2.3	1223	4 US-08-931-668-4	Sequence 4, Appli
45	33	2.3	1223	4 US-09-468-175-4	Sequence 4, Appli

RESULT 1
US-08-871-572B-3
Sequence 3, Application US/08871572B
Patent No. 6287853
GENERAL INFORMATION:
APPLICANT: Pestka, Sidney
APPLICANT: Kotenko, Serguei
APPLICANT: Soh, Jaemog
APPLICANT: Donnelly, Robert
APPLICANT: Mariano, Thomas
APPLICANT: Cook, Jeffrey
APPLICANT: Emmanuel, Stuart
APPLICANT: Schwartz, Barbara
TITLE OF INVENTION: Accessory Factor for Interferon Gamma
TITLE OF INVENTION: and Its Receptor
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard R. Muccino
STREET: 758 Springfield Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871,572B
FILING DATE: 9-JUNE-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Muccino, Richard R.
REGISTRATION NUMBER: 32,538
REFERENCE/DOCKET NUMBER: UMD1-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 273-4988
TELEFAX: (908) 273-4679
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2255 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-871-572B-3

Query Match 2.5%; Score 36; DB 3; Length 2255;
Best Local Similarity 100.0%; Pred. No. 2.5e-06;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1406 TGTGTTTAAAAA
DB 2209 TGTGTTTAAAAA 2244

RESULT 2

US-09-362-473-5
Sequence 5, Application US/09362473
Patent No. 6218169

GENERAL INFORMATION:

APPLICANT: Cahoon, Edgar B.
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Falco, S. Carl
APPLICANT: Morgante, Michele
APPLICANT: Rafalski, J. Antoni
APPLICANT: Hiltz, William D.
APPLICANT: Kinney, Anthony J.
TITLE OF INVENTION: Aromatic Amino Acid Catabolism Enzymes
FILE REFERENCE: BB-1197
CURRENT APPLICATION NUMBER: US/09/362,473
EARLIER FILING DATE: 1999-07-28
EARLIER APPLICATION NUMBER: 60/094,783
EARLIER FILING DATE: JULY 31, 1998
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Microsoft Office 97
SEQ ID NO 5
LENGTH: 1639
TYPE: DNA
ORGANISM: Glycine max
US-09-362-473-5

Query Match 2.4%; Score 35; DB 3; Length 1639;
Best Local Similarity 100.0%; Pred. No. 7e-06;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1407 GTTTTAAAAA
DB 1591 GTTTTAAAAA 1625

RESULT 3

US-08-702-344-26
Sequence 26, Application US/08702344
Patent No. 5723315

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavallie, Edward
APPLICANT: Racie, Lisa
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/702,344

FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.

REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-702-344-26

Query Match 2.4%; Score 34; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAA
DB 17 TTTTAAAAA 50

RESULT 4

US-09-621-976-13740
Sequence 13740, Application US/09621976
Patent No. 6639063

GENERAL INFORMATION:

APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 13740
LENGTH: 348
TYPE: DNA
ORGANISM: Homo sapiens
US-09-621-976-13740

Query Match 2.4%; Score 34; DB 4; Length 348;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAA
DB 314 TTTTAAAAA 347

RESULT 5

US-09-621-976-18573
Sequence 18573, Application US/09621976
Patent No. 6639063

GENERAL INFORMATION:

APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 18573
LENGTH: 388
TYPE: DNA
ORGANISM: Homo sapiens

US-09-621-976-18573

Query Match 2.4%; Score 34; DB 4; Length 388;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
|||||
Db 343 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 376

RESULT 6

US-09-640-173-10/c
; Sequence 10, Application US/09640173
; Patent No. 6613515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiaqichun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640,173
; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(396)
; OTHER INFORMATION: n = A,T,C or G
US-09-640-173-10

Query Match 2.4%; Score 34; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
|||||
Db 110 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 77

RESULT 7

US-09-713-550-10/c
; Sequence 10, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiaqichun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C4
; CURRENT APPLICATION NUMBER: US/09/713,550
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(396)
; OTHER INFORMATION: n = A,T,C or G
US-09-713-550-10

Query Match 2.4%; Score 34; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441

Db 110 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 77

RESULT 8

US-09-621-976-18571
; Sequence 18571, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18571
; LENGTH: 397
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18571

Query Match 2.4%; Score 34; DB 4; Length 397;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
|||||
Db 343 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 376

RESULT 9

US-08-567-882-5
; Sequence 5, Application US/08567882
; Patent No. 6512103
; GENERAL INFORMATION:
; APPLICANT: Dairaghi, Daniel J.
; APPLICANT: Hara, Takahiko
; APPLICANT: Miyajima, Atsushi
; APPLICANT: Schall, Thomas J.
; APPLICANT: Wang, Wei
; APPLICANT: Yoshimura, Akihiko
; TITLE OF INVENTION: MAMMALIAN CHEMOKINE REAGENTS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/567,882
; FILING DATE: 08-DEC-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0506
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 base pairs
; TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-567-882-5

Query Match 2.4%; Score 34; DB 4; Length 769;
Best Local Similarity 100.0%; Pred. No. 2e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAA 1441
DB 523 TTTTAAAAAAAAAAAAAAAAAAAAAAAA 556

RESULT 10
US-09-282-305-13
Sequence 13, Application US/09282305
Patent No. 6287843
GENERAL INFORMATION:

APPLICANT: Baldwin, Donald A.
APPLICANT: Briggs, Steven P.
APPLICANT: Crane, Virginia C.
TITLE OF INVENTION: Maize Histone Deacetylases And Their Uses
FILE REFERENCE: 5718-44,
CURRENT APPLICATION NUMBER: US/09/282,305
CURRENT FILING DATE: 1999-03-31
PRIOR APPLICATION NUMBER: 60/080,563
PRIOR FILING DATE: 1998-04-03
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 13
LENGTH: 1191
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: CDS
LOCATION: (63)..(971)
US-09-282-305-13

Query Match 2.4%; Score 34; DB 3; Length 1191;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAA 1441
DB 1158 TTTTAAAAAAAAAAAAAAAAAAAAAAAA 1191

RESULT 11
US-09-883-720-13
Sequence 13, Application US/09883720
Patent No. 6479629
GENERAL INFORMATION:

APPLICANT: Baldwin, Donald A.
APPLICANT: Briggs, Steven P.
APPLICANT: Crane, Virginia C.
TITLE OF INVENTION: Maize Histone Deacetylases And Their Uses
FILE REFERENCE: 5718-44,
CURRENT APPLICATION NUMBER: US/09/883,720
CURRENT FILING DATE: 2001-06-18
PRIOR APPLICATION NUMBER: 09/282,305
PRIOR FILING DATE: 1999-03-31
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 13
LENGTH: 1191
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: CDS
LOCATION: (63)..(971)
US-09-883-720-13

Query Match 2.4%; Score 34; DB 4; Length 1191;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAA 1441
DB 1158 TTTTAAAAAAAAAAAAAAAAAAAAAAAA 1191

RESULT 12
US-08-174-467-19
Sequence 19, Application US/08174467
Patent No. 5451514
GENERAL INFORMATION:

APPLICANT: BOUDET, ALAIN M.
APPLICANT: INZE, DIRK G.
APPLICANT: SCHUCH, WOLFGANG W.
TITLE OF INVENTION: MODIFICATION OF LIGNIN SYNTHESIS IN
TITLE OF INVENTION: PLANTS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/174,467

FILING DATE: 28-DEC-1993

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/874,166

FILING DATE: 27-APR-1992

ATTORNEY/AGENT INFORMATION:

NAME: KOKULIS, PAUL N.

REGISTRATION NUMBER: 16,773

REFERENCE/DOCKET NUMBER: 95563/PS36321/US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-861-3000

TELEFAX: 202-822-0944

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 1283 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA
US-08-174-467-19

Query Match 2.4%; Score 34; DB 1; Length 1283;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAA 1441
DB 1244 TTTTAAAAAAAAAAAAAAAAAAAAAAAA 1277

RESULT 13

US-08-452-071-19

Sequence 19, Application US/08452071
Patent No. 6066780
GENERAL INFORMATION:

APPLICANT: BOUDET, ALAIN M.
APPLICANT: INZE, DIRK G.
APPLICANT: SCHUCH, WOLFGANG W.
TITLE OF INVENTION: MODIFICATION OF LIGNIN SYNTHESIS IN

TITLE OF INVENTION: PLANTS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN, DAREY & CUSHMAN
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,071
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/874,166
FILING DATE: 27-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 95563/PS36321/US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1283 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-452-071-19

Query Match 2.4%; Score 34; DB 3; Length 1283;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
DB 1244 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1277

RESULT 14
US-09-614-912-191
Sequence 191, Application US/09614912
Patent No. 6677502
GENERAL INFORMATION:
APPLICANT: Allen, Steve
APPLICANT: Rafalski, Antoni
APPLICANT: Orozco, Buddy
APPLICANT: Miao, Gou-Hau
APPLICANT: Famodu, Omolayo O.
APPLICANT: Lee, Jian Ming
APPLICANT: Sakai, Hajime
APPLICANT: Weng, Zude
APPLICANT: Calmi, Perry G
APPLICANT: Anderson, Shawn
TITLE OF INVENTION: Plant Metabolism Genes
FILE REFERENCE: B81378 US NA
CURRENT APPLICATION NUMBER: US/09/614,912
CURRENT FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: 60/143,401
PRIOR FILING DATE: 1999-07-12
PRIOR APPLICATION NUMBER: 60/143,412
PRIOR FILING DATE: 1999-07-12
PRIOR APPLICATION NUMBER: 60/146,650
PRIOR FILING DATE: 1999-07-30
PRIOR APPLICATION NUMBER: 60/170,906
PRIOR FILING DATE: 1999-12-15

PRIOR APPLICATION NUMBER: 60/172,959
PRIOR FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: 60/172,946
PRIOR FILING DATE: 1999-12-21
NUMBER OF SEQ ID NOS: 204
SOFTWARE: Microsoft Office 97
SEQ ID NO 191
LENGTH: 1518
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: unsure
LOCATION: (114)
NAME/KEY: unsure
LOCATION: (123)
NAME/KEY: unsure
LOCATION: (138)
US-09-614-912-191

Query Match 2.4%; Score 34; DB 4; Length 1518;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1441
DB 1434 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1467

RESULT 15
US-08-545-196B-10
Sequence 10, Application US/08545196B
Patent No. 6080577
GENERAL INFORMATION:
APPLICANT: MEIKI, JUDITH
APPLICANT: MUNNICH, ARNOLD
TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,196B
FILING DATE: 19-OCT-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FARACI, C. J.
REGISTRATION NUMBER: 32,350
REFERENCE/DOCKET NUMBER: 2121-110P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1582 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-545-196B-10

Query Match 2.4%; Score 34; DB 3; Length 1582;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1408 TTTTAAAAAAAAAAAAAAAAAAAAAAAA 1441
Db 1485 TTTTAAAAAAAAAAAAAAAAAAAAAAAA 1518

Search completed: February 24, 2004, 03:24:16
Job time : 132.797 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 23:36:30 ; Search time 552.173 Seconds
(without alignments)
9138.208 Million cell updates/sec

Title: US-09-936-456-1
Perfect score: 1441
Sequence: 1 aaagtaacgctacagacag.....aaaaaaaaaaaaaaaa 1441

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2308684 seqs, 1750822206 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query				ID	Description	
No.	Score	Match	Length	DB			
1	1413	98.1	1421	14	US-10-187-657-2	Sequence 2, Appli	
2	1404	97.4	1420	14	US-10-187-657-4	Sequence 4, Appli	
3	1285	89.2	1474	15	US-10-264-237-1097	Sequence 1097, Ap	
c	4	392	27.2	2684	15	US-10-094-749-1195	Sequence 1195, Ap
	5	313	21.7	522	14	US-10-187-657-5	Sequence 5, Appli
6	187	13.0	232	9	US-09-783-590-9492	Sequence 9492, Ap	
7	136	9.4	346	14	US-10-187-657-3	Sequence 3, Appli	
8	119	8.3	450	14	US-10-187-657-7	Sequence 7, Appli	
9	107	7.4	439	15	US-10-027-632-91970	Sequence 91970, A	
10	59	4.1	512	14	US-10-106-698-3277	Sequence 3277, Ap	
11	56	3.9	506	14	US-10-187-657-6	Sequence 6, Appli	
12	38	2.6	1398	12	US-10-424-599-64755	Sequence 64755, A	
13	38	2.6	2165	14	US-10-116-255-18	Sequence 18, Appli	
14	38	2.6	3095	12	US-10-424-599-75074	Sequence 75074, A	
15	38	2.6	5642	14	US-10-311-455-842	Sequence 842, App	

16	38	2.6	10480	14	US-10-311-455-2174	Sequence 2174, Ap	
17	37	2.6	554	12	US-10-424-599-41481	Sequence 41481, A	
c	18	37	2484	12	US-10-424-599-8115	Sequence 8115, Ap	
19	37	2.6	367378	14	US-10-312-841-1	Sequence 1, Appli	
20	36	2.5	896	14	US-10-106-698-234	Sequence 234, App	
c	21	35	290	15	US-10-242-535A-52594	Sequence 52594, A	
c	22	35	302	12	US-10-424-599-48927	Sequence 48927, A	
c	23	35	379	12	US-10-424-599-43341	Sequence 43341, A	
c	24	35	379	12	US-10-424-599-133773	Sequence 133773, A	
c	25	35	416	12	US-10-424-599-50281	Sequence 50281, A	
c	26	35	449	12	US-10-424-599-1341	Sequence 1341, Ap	
c	27	35	450	14	US-10-198-846-13053	Sequence 13053, A	
c	28	35	450	15	US-10-131-827-8647	Sequence 8647, Ap	
c	29	35	464	15	US-10-131-827-8179	Sequence 8179, Ap	
c	30	35	478	12	US-10-424-599-22997	Sequence 22997, A	
c	31	35	481	12	US-10-424-599-120735	Sequence 120735, A	
c	32	35	482	12	US-10-424-599-96207	Sequence 96207, A	
c	33	35	494	12	US-10-424-599-17387	Sequence 17387, A	
c	34	35	498	12	US-10-424-599-17873	Sequence 17873, A	
c	35	35	509	12	US-10-424-599-107007	Sequence 107007, A	
c	36	35	529	14	US-10-029-386-3130	Sequence 3130, Ap	
c	37	35	621	12	US-10-424-599-10941	Sequence 10941, A	
c	38	35	658	12	US-10-424-599-122157	Sequence 122157, A	
c	39	35	1167	15	US-10-027-632-260991	Sequence 260991, A	
40	35	2.4	1639	9	US-09-742-954-5	Sequence 5, Appli	
41	35	2.4	1758	12	US-10-424-599-91053	Sequence 91053, A	
c	42	34	50	9	US-09-783-590-6670	Sequence 6670, Ap	
c	43	34	60	13	US-10-057-940-14	Sequence 14, Appli	
c	44	34	64	13	US-10-057-940-1	Sequence 1, Appli	
c	45	34	2.4	64	13	US-10-057-940-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-10-187-657-2
Sequence 2, Application US/10187657
Publication No. US20030068311A1
GENERAL INFORMATION:
APPLICANT: Laser, Amy K.W.
APPLICANT: Baughn, Mariah R.
APPLICANT: Azimzai, Yalda
TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
FILE REFERENCE: PV-0009 CIP
CURRENT APPLICATION NUMBER: US/10/187, 657
PRIOR FILING DATE: 2002-07-01
PRIOR APPLICATION NUMBER: PCT/US00/07817
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/139,565
PRIOR FILING DATE: 1999-06-16
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PERL Program
SEQ ID NO 2
LENGTH: 1421
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. US20030068311A1 4901066CB1
US-10-187-657-2

Query Match 98.1%; Score 1413; DB 14; Length 1421;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTAACGGCTACAGACAGTGAAGTAAGTTGCTGCGCGGCTAGAAAACTGTGCG 60
DB 9 AAAGTAACGGCTACAGACAGTGAAGTAAGTTGCTGCGCGGCTAGAAAACTGTGCG 68
QY 61 GTACCAACCCAGAGCGTTGAGAGCAGCCACCTCCAGGCTTCTTAACGAGAGGTGCA 120
DB 69 GTACCAACCCAGAGCGTTGAGAGCAGCCACCTCCAGGCTTCTTAACGAGAGGTGCA 128

```
QY 121 GGACTCAGACTTCAACCAAGCCCACTGGTCCCAAGCTTGTACGCAAGAGAGAGCCCAAGGAC 180
DB 129 GGACTCAGACTTCAACCAAGCCCACTGGTCCCAAGCTTGTACGCAAGAGAGAGCCCAAGGAC 188
QY 181 GCGCTCTCCCGCGTCCAGGAGAGCCCAAGCTTGTGCTGCTTGCCTGCCCCGCTGCTGAGC 240
DB 189 GCGCTCTCCCGCGTCCAGGAGAGCCCAAGCTTGTGCTGCTTGCCTGCCCCGCTGCTGAGC 248
QY 241 ACTCGCCGCGGTGAGAGATGACCTGTGAAACGCGTACTGCTTTTACCAGCCAGCC 300
DB 249 ACTCGCCGCGGTGAGAGATGACCTGTGAAACGCGTACTGCTTTTACCAGCCAGCC 308
QY 301 CGGATGCGGAGAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCA 360
DB 309 CGGATGCGGAGAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCA 368
QY 361 GCAGCAAGCTTCTGCTCATCTTGCAGGAGATCCGTGGCCACTGCGCTGTTGTTG 420
DB 369 GCAGCAAGCTTCTGCTCATCTTGCAGGAGATCCGTGGCCACTGCGCTGTTGTTG 428
QY 421 GTGAGAGTTCTTCTCAGTCTGTTCATAGGCGAGAAATTGTGCTGTGCACTTCACTGCA 480
DB 429 GTGAGAGTTCTTCTCAGTCTGTTCATAGGCGAGAAATTGTGCTGTGCACTTCACTGCA 488
QY 481 GAATGTTGTTGGTACAGTGAACCAACACATCTTCAAAAGCCTTCAAGCGAGCGCGC 540
DB 489 GAATGTTGTTGGTACAGTGAACCAACACATCTTCAAAAGCCTTCAAGCGAGCGCGC 548
QY 541 GTTACAGCCCGTGTGCTGTCTGCTGCTGCTGCTGAGGCGATTAATTAATTAATTAATTA 600
DB 549 GTTACAGCCCGTGTGCTGTCTGCTGCTGCTGAGGCGATTAATTAATTAATTAATTAATTA 608
QY 601 ACCCCAGTGCATCAGTGAACGAGACCACTTGAATTAATTAATTAATTAATTAATTAATTA 660
DB 609 ACCCCAGTGCATCAGTGAACGAGACCACTTGAATTAATTAATTAATTAATTAATTAATTA 668
QY 661 AAAGAGAAATTAACGCGCGGAGTAACGCGACGCACTGAGAAAGGGGCTCCGAGACCAAGT 720
DB 669 AAAGAGAAATTAACGCGCGGAGTAACGCGACGCACTGAGAAAGGGGCTCCGAGACCAAGT 728
QY 721 CTCTACCTGGGAGAAAGTTCACACCGAGTAGCCCTTGGCGCTGTACCAACAGTACCAAC 780
DB 729 CTCTACCTGGGAGAAAGTTCACACCGAGTAGCCCTTGGCGCTGTACCAACAGTACCAAC 788
QY 781 CTGGCGGAGACACTACGCTTGGCGAGCACTATGGGTGCTTCTGCTTCTGCTCTCTCC 840
DB 789 CTGGCGGAGACACTACGCTTGGCGAGCACTATGGGTGCTTCTGCTTCTGCTCTCTCC 848
QY 841 AAGCTGCTGCTTCCACGCGCGCGCGCTGTAAGGAGGCTGAGCACTGTGACCAACCGGA 900
DB 849 AAGCTGCTGCTTCCACGCGCGCGCGCTGTAAGGAGGCTGAGCACTGTGACCAACCGGA 908
QY 901 GCCTTGGCGCTTCTGCGGAGTCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 960
DB 909 GCCTTGGCGCTTCTGCGGAGTCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 968
QY 961 CTCGCGCTAGGCTCTCCGCGCTCAACCACTCACTAGTACGCGCGCGCTTCTGGGTCAAGCTG 1020
DB 969 CTCGCGCTAGGCTCTCCGCGCTCAACCACTCACTAGTACGCGCGCGCTTCTGGGTCAAGCTG 1028
QY 1021 GCAACCGGCGTCTGCTGCTTCTGCTGAGGAGGCGCGTGTAGTCTCCAGTATGTTGG 1080
DB 1029 GCAACCGGCGTCTGCTGCTTCTGCTGAGGAGGCGCGTGTAGTCTCCAGTATGTTGG 1088
QY 1081 CCCAGCGCTCTTGCACACCTTCTGAGCAAAAGCGCCAAAGACTGCAAGAGAGAGGG 1140
DB 1089 CCCAGCGCTCTTGCACACCTTCTGAGCAAAAGCGCCAAAGACTGCAAGAGAGAGGG 1148
QY 1141 GGCTCAGCTCTTATCTCGGCGAGACCACTGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1149 GGCTCAGCTCTTATCTCGGCGAGACCACTGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1208
```

```
QY 1201 TGTATCACCATAACCTGTGAGGAGGAGACCCCAATCTGGAATCTCTTCCCGCTTGGAGAT 1260
DB 1209 TGTATCACCATAACCTGTGAGGAGGAGACCCCAATCTGGAATCTCTTCCCGCTTGGAGAT 1268
QY 1261 CGCAGGCGGAGAGAGAGTCCCGCGAGGCTTGGCGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
DB 1269 CGCAGGCGGAGAGAGAGTCCCGCGAGGCTTGGCGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1328
QY 1321 AGCGCTGCTGGCGGAGAGGCTTGGAGATCCGAGAGGACCAAGGAAAGTCTCTGGGGCGA 1380
DB 1329 AGCGCTGCTGGCGGAGAGGCTTGGAGATCCGAGAGGACCAAGGAAAGTCTCTGGGGCGA 1388
QY 1381 TCTGTAATTAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 1413
DB 1389 TCTGTAATTAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 1421
```

RESULT 2

```
US-10-187-657-4
; Sequence 4, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 1420
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030068311A1 3221661CA2
US-10-187-657-4
```

```
Query Match 97.4%; Score 1404; DB 14; Length 1420;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAGTAACGGCTACAGACAGTGAAGAAATAGTTTGGCTCGCGCGGCTAGAAAACCTGTG 60
DB 9 AAAGTAACGGCTACAGACAGTGAAGAAATAGTTTGGCTCGCGCGGCTAGAAAACCTGTG 68
QY 61 GTACCAACCCCAAGAGCGTTGAGAGAGAGCCCACTCCAGGCTTCTTAACGAGAGAGTGA 120
DB 69 GTACCAACCCCAAGAGCGTTGAGAGAGAGCCCACTCCAGGCTTCTTAACGAGAGAGTGA 128
QY 121 GGACTCAGACTTCAACCAAGCCCACTCGTCCAGGCTTGTACGCAAAAGAGAGAGAGAGAG 180
DB 129 GGACTCAGACTTCAACCAAGCCCACTCGTCCAGGCTTGTACGCAAAAGAGAGAGAGAGAG 188
QY 181 GCGCTCTCCCGCGTCCAGGAGAGCCCAAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
DB 189 GCGCTCTCCCGCGTCCAGGAGAGCCCAAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 248
QY 241 ACTCGCCGCGGTGAGAGATGACCTGTGAAACGCGTACTGCTTTTACCAGCCAGCC 300
DB 249 ACTCGCCGCGGTGAGAGATGACCTGTGAAACGCGTACTGCTTTTACCAGCCAGCC 308
QY 301 CGGATGCGGAGAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCA 360
DB 309 CGGATGCGGAGAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCA 368
QY 361 GCAGCAAGCTTCTGCTCATCTTGCAGGAGATCCGTGGCCACTGCGCTGTTGTTG 420
```

Db	369	GCAGCAAGCTTCCTGCTCATCTTGCCGGGGATCCGTGACCACCTCGCGCTGTTTGGTTG	428
QY	421	GTGAGAGTCTTCTCAGTCTGTTCATATAGCGCAGAAATGTGTGCTGTGCACCTTCAGTGCA	480
Db	429	GTGAGAGTCTTCTCAGTCTGTTCATATAGCGCAGAAATGTGTGCTGTGCACCTTCAGTGCA	488
QY	481	GAATGTTCTGGGTACAGTGAACACCAACACATCCTACAAAGCCTTCAGCGCAGCGCGC	540
Db	489	GAATGTTCTGGGTACAGTGAACACCAACACATCCTACAAAGCCTTCAGCGCAGCGCGC	548
QY	541	GTTACAGCCCGTGTGCGTCTGCTGTGGGCTTGAGGGCATTAAATTAACCTCACAGGG	600
Db	549	GTTACAGCCCGTGTGCGTCTGCTGTGGGCTTGAGGGCATTAAATTAACCTCACAGGG	608
QY	601	AACCCAGTGATCAGCTGAACGAGAACCATTTGACTACAAAGCAGAGCTTCACTGGCGTCTG	660
Db	609	AACCCAGTGATCAGCTGAACGAGAACCATTTGACTACAAAGCAGAGCTTCACTGGCGTCTG	668
QY	661	AAAGAGAATTACGCCCGCAGATACGCGAACGCACCTGGAGAAAGGGCTGCCGGAACCACTG	720
Db	669	AAAGAGAATTACGCCCGCAGATACGCGAACGCACCTGGAGAAAGGGCTGCCGGAACCACTG	728
QY	721	CTCTACCTGGCGGAGAGATTACACACCGAGTAGCCCTTGCGGCTGTATCCACCAAGTACCAC	780
Db	729	CTCTACCTGGCGGAGAGATTACACACCGAGTAGCCCTTGCGGCTGTATCCACCAAGTACCAC	788
QY	781	CTGGCGGACACTACGCGCTGGCCACAGCTATGGGTGCGCTTCTGCTTCTGGCTCCTCTCC	840
Db	789	CTGGCGGACACTACGCGCTGGCCACAGCTATGGGTGCGCTTCTGCTTCTGGCTCCTCTCC	848
QY	841	AACGTGCTGCTCTCCACGCGCGGCCCGCTCTACGGAAGGCTGGCACTGCTGACCAACCGGA	900
Db	849	AACGTGCTGCTCTCCACGCGCGGCCCGCTCTACGGAAGGCTGGCACTGCTGACCAACCGGA	908
QY	901	GCCTTGCGGCTCTTCGGGGTCTTCGCTTGGCCTCCATCTTACGCTGCGCTCTGCCCG	960
Db	909	GCCTTGCGGCTCTTCGGGGTCTTCGCTTGGCCTCCATCTTACGCTGCGCTCTGCCCG	968
QY	961	CTCCGCTAGGCTCTCCGCGCTCACCACTCAGTACGCGCGCGCTTCTGGGTACCGCTG	1020
Db	969	CTCCGCTAGGCTCTCCGCGCTCACCACTCAGTACGCGCGCGCTTCTGGGTACCGCTG	1028
QY	1021	GCAACCGGCGTCTCTGTGCTCTTCTCCCTCGAGGGGCGGTGTGAGTCTCCAGTATGTCG	1080
Db	1029	GCAACCGGCGTCTCTGTGCTCTTCTCCCTCGAGGGGCGGTGTGAGTCTCCAGTATGTCG	1088
QY	1081	CCGACGCGCTCTTGCAACCCCTTCTGGAACCAAGGCCCAAGGACTGCAGCCAGAGAGGG	1140
Db	1089	CCGACGCGCTCTTGCAACCCCTTCTGGAACCAAGGCCCAAGGACTGCAGCCAGAGAGGG	1148
QY	1141	GGCTCACTCTTATCTCTCGGGAACCACTGCACAAGCAGCGCTCTCCAGACTTAAAA	1200
Db	1149	GGCTCACTCTTATCTCTCGGGAACCACTGCACAAGCAGCGCTCTCCAGACTTAAAA	1208
QY	1201	TGTATCAACCACTAAGCTGTGAGGGGGAACCAATCTGAGCTCTTCCCGCTTGGGACAT	1260
Db	1209	TGTATCAACCACTAAGCTGTGAGGGGGAACCAATCTGAGCTCTTCCCGCTTGGGACAT	1268
QY	1261	CGACGGCCGGGAAGCAGTCCCGCCAGAGCCTTGAGCCAGAGAGCTCCAGAAAGGCACTG	1320
Db	1269	CGACGGCCGGGAAGCAGTCCCGCCAGAGCCTTGAGCCAGAGAGCTCCAGAAAGGCACTG	1328
QY	1321	AGCGCTGTGGCGGAGGCTTCGACATCCGCAAGCACCAAGGAAAGTCTCTGGGCGA	1380
Db	1329	AGCGCTGTGGCGGAGGCTTCGACATCCGCAAGCACCAAGGAAAGTCTCTGGGCGA	1388
QY	1381	TCTGTAAATAAACCTTTTCTTT 1404	
Db	1389	TCTGTAAATAAACCTTTTCTTT 1412	

```

US-10-264-237-1097
/ Sequence 1097, Application US/10264237
/ Publication No. US20040009491A1
/ GENERAL INFORMATION:
/ APPLICANT: Birse et al.
/ TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
/ FILE REFERENCE: PA131PI
/ CURRENT APPLICATION NUMBER: US/10/264,237
/ CURRENT FILING DATE: 2002-10-04
/ PRIOR APPLICATION NUMBER: PCT/US01/16450
/ PRIOR FILING DATE: 2001-05-18
/ PRIOR APPLICATION NUMBER: US 60/205,515
/ PRIOR FILING DATE: 2000-05-19
/ NUMBER OF SEQ ID NOS: 2876
/ SOFTWARE: PatentIn Ver. 3.1
/ SEQ ID NO 1097
/ LENGTH: 1474
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-264-237-1097

```

Query Match	89.2%	Score 1285	DB 15	Length 1474
Best Local Similarity	99.8%	Pred. No. 0		
Matches 1435	Conservative 0	Mismatches 3	Indels 0	Gaps 0
QY	1	AAAGTACGGCTACAGACAGTGAAGAAATAGTTCCGCTCGCCGGCTAGAAAAA	CTGTGCG	60
Db	27	AAAGTACGGCTACAGACAGTGAAGAAATAGTTCCGCTCGCCGGCTAGAAAAA	CTGTGCG	85
QY	61	GTACCAACCCCGAGAGCGTTGAGAGCAGCCCACTCCACGCTTCTTAA	CGGAGGTGCA	120
Db	87	GTACCAACCCCGAGAGCGTTGAGAGCAGCCCACTCCACGCTTCTTAA	CGGAGGTGCA	145
QY	121	GGACTCAGACTTCACCAAGCCCACTCGGTCCAGCCTTGTACGAAAGAGAC	CGCAAGGAC	180
Db	147	GGACTCAGACTTCACCAAGCCCACTCGGTCCAGCCTTGTACGAAAGAGAC	CGCAAGGAC	206
QY	181	GCGCTCTCCCGCGTCCAGGACGCCCAAGCTTGCTGGCTTGCTGCGCGCT	CGCTGACAGC	240
Db	207	GCGCTCTCCCGCGTCCAGGACGCCCAAGCTTGCTGGCTTGCTGCGCGCT	CGCTGACAGC	266
QY	241	ACTCGCCCGGCGTGACAGATGACCCTGTGGAACGGCGTACTGCTTTT	TAAACCCAGCCC	300
Db	267	ACTCGCCCGGCGTGACAGATGACCCTGTGGAACGGCGTACTGCTTTT	TAAACCCAGCCC	326
QY	301	CGGCATGCCGAGGCTTCAAGCGTTCCACTGCTCATCGTTATTCTAG	GTGTTTTGGCTCTA	360
Db	327	CGGCATGCCGAGGCTTCAAGCGTTCCACTGCTCATCGTTATTCTAG	GTGTTTTGGCTCTA	386
QY	361	GCAGCAAGCTTCCTGCTCATCTTGCCGGGGATCCGTTGACCACTCGCGCT	GGTTTTGGTTG	420
Db	387	GCAGCAAGCTTCCTGCTCATCTTGCCGGGGATCCGTTGACCACTCGCGCT	GGTTTTGGTTG	446
QY	421	GTGAGAGTTCTTCTCAGTCTGTTTCATATGGCGCAGAAATGTGCTGTG	CACTTCAGTGCA	480
Db	447	GTGAGAGTTCTTCTCAGTCTGTTTCATATGGCGCAGAAATGTGCTGTG	CACTTCAGTGCA	506
QY	481	GAATGGTTCGTGGGTACAGTGAACACCAACACATCCTAACAAAGCCTT	CAGCGCAGCGCC	540
Db	507	GAATGGTTCGTGGGTACAGTGAACACCAACACATCCTAACAAAGCCTT	CAGCGCAGCGCC	566
QY	541	GTATACAGCCCGTGTGCGTCTGCTCGTGGGCTTGAGGGCATTAATAT	TACACTCACAGGG	600
Db	567	GTATACAGCCCGTGTGCGTCTGCTCGTGGGCTTGAGGGCATTAATAT	TACACTCACAGGG	626
QY	601	ACCCAGTGATCAGCTGAACGAGACCATTTGACTACAAAGAGCAGTTC	CACTTGGCGTCTG	660
Db	627	ACCCAGTGATCAGCTGAACGAGACCATTTGACTACAAAGAGCAGTTC	CACTTGGCGTCTG	686
QY	661	AAAGAGAATTACGCCCGGAGAGTACGGAACGCACTGGAGAAAGGGGCT	CGCCGAGCCAGTG	720
Db	687	AAAGAGAATTACGCCCGGAGAGTACGGAACGCACTGGAGAAAGGGGCT	CGCCGAGCCAGTG	746

```
QY 721 CTCTACCTGGCGAGAGTTTACACCGAGTAGCCCTTGCGGCTGTACCACTAGTACCAC 780
Db 747 CTCTACCTGGCGAGAGTTTACACCGAGTAGCCCTTGCGGCTGTACCACTAGTACCAC 806
QY 781 CTGCGGGAACACTACGCTTCCGCGGCAAGTATGGGTGGCGTTCTTCTTCTGCTCTCC 840
Db 807 CTGCGGGAACACTACGCTTCCGCGGCAAGTATGGGTGGCGTTCTTCTTCTGCTCTCC 866
QY 841 AACGTGCTGCTCTCCAGCGCGGCGGCTCTACGAGGCGCTGCACTGTGACCAACCGGA 900
Db 867 AACGTGCTGCTCTCCAGCGCGGCGGCTCTACGAGGCGCTGCACTGTGACCAACCGGA 926
QY 901 GCCTTCGCGCTCTTCCGCGGCTTCCGCTTGGCTTCCATCTTACGCTGCGCTCTGCCCG 960
Db 927 GCCTTCGCGCTCTTCCGCGGCTTCCGCTTGGCTTCCATCTTACGCTGCGCTCTGCCCG 986
QY 961 CTCGCGCTAGGCTCTCTCGCGGCTCAGCACTCAGTACGGCGCGCTTCTGGGTCAAGCTG 1020
Db 987 CTCGCGCTAGGCTCTCTCGCGGCTCAGCACTCAGTACGGCGCGCTTCTGGGTCAAGCTG 1046
QY 1021 GCAACCGGCGTCTGTGCTCTTCTCTGAGAGGGCGGTGTAGTCTCCAGTATGTTCCG 1080
Db 1047 GCAACCGGCGTCTGTGCTCTTCTCTGAGAGGGCGGTGTAGTCTCCAGTATGTTCCG 1106
QY 1081 CCCAGCGCTCTTGCACCCCTTCTGACCAAGCGGCAAGGACTGCAAGGAGAGAGAGG 1140
Db 1107 CCCAGCGCTCTTGCACCCCTTCTGACCAAGCGGCAAGGACTGCAAGGAGAGAGAGG 1166
QY 1141 GGCTCAGCTCTTATCTCGGCGGCACTGCAAGGAGAGGCGCTCTCCAGCTTAA 1200
Db 1167 GGCTCAGCTCTTATCTCGGCGGCACTGCAAGGAGAGGCGCTCTCCAGCTTAA 1226
QY 1201 TGTATCACTAATACCTGTGAGGGGGAACCAATCTGACTCTTCCCGCTTGGGACAT 1260
Db 1227 TGTATCACTAATACCTGTGAGGGGGAACCAATCTGACTCTTCCCGCTTGGGACAT 1286
QY 1261 CGCAGCGCGGGAAGCAGTGGCGGCGGCAAGCTTGGCGGAGAGAGCTTCCAGGAGGCACTG 1320
Db 1287 CGCAGCGCGGGAAGCAGTGGCGGCGGCAAGCTTGGCGGAGAGAGCTTCCAGGAGGCACTG 1346
QY 1321 AGCGCTGTGGCGGAGGCTTCCAGTCCGAGGCAACAGGAAAGTCTCTGCGGCGGA 1380
Db 1347 AGCGCTGTGGCGGAGGCTTCCAGTCCGAGGCAACAGGAAAGTCTCTGCGGCGGA 1406
QY 1381 TCTGTAATATAACCTTTTCTTTTCTTTTAA 1438
Db 1407 TCTGTAATATAACCTTTTCTTTTCTTTTAA 1464
```

RESULT 4

```
US-10-094-749-1195/c
; Sequence 1195, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIB, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
```

```
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1195
; LENGTH: 2684
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-094-749-1195

Query Match      27.2%; Score 392; DB 15; Length 2684;
Best Local Similarity 100.0%; Pred. No. 3e-181;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1027 GCGCTCCTGTGCTCTTCTCTCGAGGGGCGGTGTAGTCTCCAGTATGTTGGGCCAGC 1086
Db 2343 GCGCTCCTGTGCTCTTCTCTCGAGGGGCGGTGTAGTCTCCAGTATGTTGGGCCAGC 2284
QY 1087 GCTCTTGCACCTTCTTGCACCAAGCGCCAGGACTGCAAGCAGAGAGAGAGGGGCTCA 1146
Db 2283 GCTCTTGCACCTTCTTGCACCAAGCGCCAGGACTGCAAGCAGAGAGAGAGGGGCTCA 2224
QY 1147 CCTCTATCTCTGCGGAGCCCACTGCAAGAGGCGCTCTCCAGACTTAAATGTATC 1206
Db 2223 CCTCTATCTCTGCGGAGCCCACTGCAAGAGGCGCTCTCCAGACTTAAATGTATC 2164
QY 1207 ACCACTAACCCTGTGAGGGGGAACCAATCTGACTCTTCCCGCTTGGGACATGCGAGG 1266
Db 2163 ACCACTAACCCTGTGAGGGGGAACCAATCTGACTCTTCCCGCTTGGGACATGCGAGG 2104
QY 1267 CCGGGAAGCAGTGGCGGCGGCAAGCTTGGCGGAGAGAGTCCAGGAGGCACTGAGGCT 1326
Db 2103 CCGGGAAGCAGTGGCGGCGGCAAGCTTGGCGGAGAGAGTCCAGGAGGCACTGAGGCT 2044
QY 1327 GCTGGCGGAGGCTTCCAGATCCGAGGCAACAGGAAAGTCTCTGCGGCGATCTGTA 1386
Db 2043 GCTGGCGGAGGCTTCCAGATCCGAGGCAACAGGAAAGTCTCTGCGGCGATCTGTA 1984
QY 1387 AATAAACCTTTTCTTTTCTTTTAA 1418
Db 1983 AATAAACCTTTTCTTTTCTTTTAA 1952
```

RESULT 5

```
US-10-187-657-5
; Sequence 5, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Lassek, Amy R.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 522
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030068311A1 3221661R6
```

FEATURE:
NAME/KEY: unsure
LOCATION: 455, 480, 483
OTHER INFORMATION: a, t, c, g, or other
US-10-187-657-5

Query Match 21.7%; Score 313; DB 14; Length 522;
Best Local Similarity 100.0%; Pred. No. 1.4e-142;
Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTAACGGCTACAGACAGTGAATAAGTTTCGCTCGCCGGCTAGAAAACTCTGTG 60
DB 9 AAAGTAACGGCTACAGACAGTGAATAAGTTTCGCTCGCCGGCTAGAAAACTCTGTG 68
QY 61 GTACCAACCCCAAGAGCGCTTGAGAGAGAGCCCACTCCAGCTTCTTAACGAGAGGTGCA 120
DB 69 GTACCAACCCCAAGAGCGCTTGAGAGAGAGCCCACTCCAGCTTCTTAACGAGAGGTGCA 128
QY 121 GGAAGTCAAGCTTCAACAG 180
DB 129 GGAAGTCAAGCTTCAACAG 188
QY 181 GCGCTCTCCCGCGCTGACAG 240
DB 189 GCGCTCTCCCGCGCTGACAG 248
QY 241 ACTCGGCGCGCGCTGACAG 300
DB 249 ACTCGGCGCGCGCTGACAG 308
QY 301 CGGCATGCGCGCAG 313
DB 309 CGGCATGCGCGCAG 321

RESULT 6

US-09-783-590-9492

; Sequence 9492, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Ruben, Craig A.
; APPLICANT: Rosen, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16,2C1
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: US/09/783,590
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 9492
; LENGTH: 232
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (202)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (206)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (229)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-9492

Query Match 13.0%; Score 187; DB 9; Length 232;
Best Local Similarity 100.0%; Pred. No. 5.6e-81;

Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTAACGGCTACAGACAGTGAATAAGTTTCGCTCGCCGGCTAGAAAACTCTGTG 60
DB 15 AAAGTAACGGCTACAGACAGTGAATAAGTTTCGCTCGCCGGCTAGAAAACTCTGTG 74
QY 61 GTACCAACCCCAAGAGCGCTTGAG 120
DB 75 GTACCAACCCCAAGAGCGCTTGAG 134
QY 121 GGAAGTCAAGCTTCAACAG 180
DB 135 GGAAGTCAAGCTTCAACAG 194
QY 181 GCGCTCT 187
DB 195 GCGCTCT 201

RESULT 7

US-10-187-657-3

; Sequence 3, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Laese, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1993-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 3
; LENGTH: 346
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030068311A1 1752794F6
; NAME/KEY: unsure
; LOCATION: 91, 188, 206, 291, 337
; OTHER INFORMATION: a, t, c, g, or other
US-10-187-657-3

Query Match 9.4%; Score 136; DB 14; Length 346;
Best Local Similarity 99.5%; Pred. No. 4.5e-56;
Matches 186; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 927 CTGGGCTCCATCTCTAGCGTGCCTCTGCGCGCTCCGCTAGGCTCCTCCGCGCTAC 986
DB 1 CTGGGCTCCATCTCTAGCGTGCCTCTGCGCGCTCCGCTAGGCTCCTCCGCGCTAC 60
QY 987 CACTAGTACGGCGCGCGCTTCTGGGTACCGCTGGCAACCGCGCTCCTGCTTCTCT 1046
DB 61 CACTAGTACGGCGCGCGCTTCTGGGTACCGCTGGCAACCGCGCTCCTGCTTCTCT 120
QY 1047 CGAGGGCGCGGTGAGTCTCCAGTATGTTGGCCCAAGCGCTTTCGACCCCTTGGGA 1106
DB 121 CGAGGGCGCGGTGAGTCTCCAGTATGTTGGCCCAAGCGCTTTCGACCCCTTGGGA 180
QY 1107 CCAAGC 1113
DB 181 CCAAGC 187

RESULT 8
US-10-187-657-7


```
/ Sequence 7, Application US/10187657
/ Publication No. US20030068311A1
/ GENERAL INFORMATION:
/ APPLICANT: Lasek, Amy K.W.
/ APPLICANT: Baughn, Mariah R.
/ APPLICANT: Azimzai, Yalda
/ TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
/ FILE REFERENCE: PV-0009 CIP
/ CURRENT APPLICATION NUMBER: US/10/187,657
/ PRIOR FILING DATE: 2002-07-01
/ PRIOR APPLICATION NUMBER: PCT/US00/07817
/ PRIOR FILING DATE: 2000-03-22
/ PRIOR APPLICATION NUMBER: 60/139,565
/ PRIOR FILING DATE: 1999-06-16
/ NUMBER OF SEQ ID NOS: 9
/ SOFTWARE: PERL Program
/ SEQ ID NO 7
/ LENGTH: 450
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No. US20030068311A1 SEQID4510D1
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: 140, 165, 203, 237, 269, 289, 307, 338, 344, 347, 354, 390, 430, 445
/ OTHER INFORMATION: a, t, c, g, or other
/ US-10-187-657-7

Query Match
Best Local Similarity 8.3%; Score 119; DB 14; Length 450;
Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CTGAACGAGACCATGTACTACACGAGAGTCACTGGCGCTCTGAAAGAGATTACGCC 675
DB 21 CTGAACGAGACCATGTACTACACGAGAGTCACTGGCGCTCTGAAAGAGATTACGCC 80

QY 676 GCGGAGTACCGGAGACGCACTGGAGAGGGGCTGCCGGAACCACTGCTCTACCTGGCGGA 734
DB 81 GCGGAGTACCGGAGACGCACTGGAGAGGGGCTGCCGGAACCACTGCTCTACCTGGCGGA 139

RESULT 9
/ US-10-027-632-91970
/ Sequence 91970, Application US/10027632
/ Publication No. US20030204075A9
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ PRIOR FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 91970
/ LENGTH: 439
/ TYPE: DNA
/ ORGANISM: Human
```

```
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)..(439)
/ OTHER INFORMATION: n = A,T,C or G
/ US-10-027-632-91970

Query Match
Best Local Similarity 7.4%; Score 107; DB 15; Length 439;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 892 ACCACCGAGAGCTTCGCGCTCTTCGGGCTCTTCGCTTGCCCTCATCTTAGCGTCCG 951
DB 214 ACCACCGAGAGCTTCGCGCTCTTCGGGCTCTTCGCTTGCCCTCATCTTAGCGTCCG 273

QY 952 CTCTGCCCCGCTCCGCTAGGCTCCCTCCGCGCTACCACTCAGTACGG 998
DB 274 CTCTGCCCCGCTCCGCTAGGCTCCCTCCGCGCTACCACTCAGTACGG 320

RESULT 10
/ US-10-106-698-3277
/ Sequence 3277, Application US/10106698
/ Publication No. US20030109690A1
/ GENERAL INFORMATION:
/ APPLICANT: Ruben et al.
/ TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypep
/ FILE REFERENCE: PA005P1
/ CURRENT APPLICATION NUMBER: US/10/106,698
/ PRIOR FILING DATE: 2002-03-27
/ PRIOR APPLICATION NUMBER: PCT/US00/26524
/ PRIOR FILING DATE: 2000-09-28
/ PRIOR APPLICATION NUMBER: US 60/157,137
/ PRIOR FILING DATE: 1999-09-29
/ PRIOR APPLICATION NUMBER: US 60/163,280
/ PRIOR FILING DATE: 1999-11-03
/ NUMBER OF SEQ ID NOS: 8564
/ SOFTWARE: PatentIn Ver. 3.0
/ SEQ ID NO 3277
/ LENGTH: 512
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (392)..(392)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc_feature
/ LOCATION: (453)..(453)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc_feature
/ LOCATION: (461)..(461)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc_feature
/ LOCATION: (493)..(493)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc_feature
/ LOCATION: (499)..(499)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc_feature
/ LOCATION: (509)..(509)
/ OTHER INFORMATION: n equals a,t,g, or c
/ US-10-106-698-3277

Query Match
Best Local Similarity 4.1%; Score 59; DB 14; Length 512;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 GCGCTGTTTGGTTGGTGAGAGTCTTCTCAGTCTGTTCTATAGCGGAGAAATTGTGG 463
DB 220 GCGCTGTTTGGTTGGTGAGAGTCTTCTCAGTCTGTTCTATAGCGGAGAAATTGTGG 278

RESULT 11
/ US-10-187-657-6
```

```
/ Sequence 6, Application US/10187657
/ Publication No. US20030068311A1
/ GENERAL INFORMATION:
/ APPLICANT: Lasek, Amy K.W.
/ APPLICANT: Baughn, Mariah R.
/ TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
/ FILE REFERENCE: PV-0009 CIP
/ CURRENT APPLICATION NUMBER: US/10/187,657
/ PRIOR FILING DATE: 2002-07-01
/ PRIOR APPLICATION NUMBER: PCT/US00/07817
/ PRIOR FILING DATE: 2000-03-22
/ PRIOR APPLICATION NUMBER: 60/139,565
/ PRIOR FILING DATE: 1999-06-16
/ NUMBER OF SEQ ID NOS: 9
/ SOFTWARE: PERL Program
/ SEQ ID NO 6
/ LENGTH: 506
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No. US20030068311A1 SBOA03652D1
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: 87, 89, 93, 95, 124, 178, 227, 253, 262, 279, 285, 287, 293-294, 298,
/ LOCATION: 300, 303, 312, 319-321, 323, 325, 327, 332-333, 336, 343-344, 349, 377, 39
/ LOCATION: 405, 409-410, 412, 415, 420, 423, 429, 432, 436-437, 444-445, 448, 452, 45
/ LOCATION: 466, 469-471, 479, 483, 485, 492, 494, 499, 501, 504
/ OTHER INFORMATION: a, t, c, g, or other
/ US-10-187-657-6

Query Match          3.9%; Score 56; DB 14; Length 506;
Best Local Similarity 100.0%; Pred. No. 5.4e-17;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 479 CAGATGTTGTCGTGGTACAGTGAACACACACATCCTACAAAGCCTTCAGCGCA 534
DB 31 CAGATGTTGTCGTGGTACAGTGAACACACACACATCCTACAAAGCCTTCAGCGCA 86

RESULT 12
US-10-424-599-64755
/ Sequence 64755, Application US/10424599
/ Publication No. US20040031072A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa Thomas J
/ APPLICANT: Kovalic David K
/ APPLICANT: Zhou Yihua
/ APPLICANT: Cao Yongwei
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 64755
/ LENGTH: 1398
/ TYPE: DNA
/ ORGANISM: Glycine max
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_29487C.1
/ US-10-424-599-64755

Query Match          2.6%; Score 38; DB 12; Length 1398;
Best Local Similarity 100.0%; Pred. No. 3.1e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1404 TTGTTTAAAAA
DB 1325 TTGTTTAAAAA 1362
```

```
RESULT 13
US-10-116-255-18
/ Sequence 18, Application US/10116255
/ Publication No. US20030036646A1
/ GENERAL INFORMATION:
/ APPLICANT: Ni et al.
/ TITLE OF INVENTION: Uncoupling Protein Polynucleotides, Polypeptides, and
/ TITLE OF INVENTION: Antibodies
/ FILE REFERENCE: PT009PI
/ CURRENT APPLICATION NUMBER: US/10/116,255
/ PRIOR FILING DATE: 2002-04-05
/ PRIOR APPLICATION NUMBER: 09/685,897
/ PRIOR FILING DATE: 2000-10-11
/ PRIOR APPLICATION NUMBER: PCT/US00/09534
/ PRIOR FILING DATE: 2000-04-06
/ PRIOR APPLICATION NUMBER: 60/128,701
/ PRIOR FILING DATE: 1999-04-09
/ PRIOR APPLICATION NUMBER: 60/142,821
/ PRIOR FILING DATE: 1999-07-08
/ PRIOR APPLICATION NUMBER: 60/149,448
/ PRIOR FILING DATE: 1999-08-18
/ PRIOR APPLICATION NUMBER: 60/164,751
/ PRIOR FILING DATE: 1999-11-12
/ NUMBER OF SEQ ID NOS: 66
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 18
/ LENGTH: 2165
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-10-116-255-18

Query Match          2.6%; Score 38; DB 14; Length 2165;
Best Local Similarity 100.0%; Pred. No. 3e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1404 TTGTTTAAAAA
DB 2128 TTGTTTAAAAA 2165

RESULT 14
US-10-424-599-75074
/ Sequence 75074, Application US/10424599
/ Publication No. US20040031072A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa Thomas J
/ APPLICANT: Kovalic David K
/ APPLICANT: Zhou Yihua
/ APPLICANT: Cao Yongwei
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 75074
/ LENGTH: 3095
/ TYPE: DNA
/ ORGANISM: Glycine max
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_38805C.1
/ US-10-424-599-75074
```

```
Query Match          2.6%; Score 38; DB 12; Length 3095;
Best Local Similarity 100.0%; Pred. No. 2.9e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1404 TTGTTTAAAAA
DB 3047 TTGTTTAAAAA 3084

RESULT 15
```

```
US-10-311-455-842
; Sequence 842, Application US/10311455
; Publication No: US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 842
; LENGTH: 5642
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-842

Query Match      2.6%; Score 38; DB 14; Length 5642;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1404 TTTGTTTTTTAAAAA
DB      1374 TTTGTTTTTTAAAAA
                        1441
                        1411

Search completed: February 24, 2004, 03:36:57
Job time : 557.173 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 23:20:25 ; Search time 3666.04 Seconds
(without alignments)
11737.841 Million cell updates/sec

Title: US-09-936-456-1
Perfect score: 1441
Sequence: 1 aaagtaacgctacagacag.....aaaaaaaaaaaaaaaa 1441

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estlin: *
4: em_estmu: *
5: em_estrov: *
6: em_esttpl: *
7: em_estro: *
8: em_htc: *
9: gb_est1: *
10: gb_est2: *
11: gb_htc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: em_gss_hum: *
18: em_gss_inv: *
19: em_gss_pln: *
20: em_gss_vrt: *
21: em_gss_fun: *
22: em_gss_mam: *
23: em_gss_mus: *
24: em_gss_pro: *
25: em_gss_rpd: *
26: em_gss_phg: *
27: em_gss_vrt: *
28: gb_gss1: *
29: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	528	36.6	537	9	AI821606	AI821606 nk08a11.x
C 2	515	35.7	515	9	AI826629	AI826629 wk35e04.x
C 3	510	35.4	510	9	AI304327	AI304327 qo57h06.x
C 4	446	31.0	446	9	AI984141	AI984141 wu21c02.x

C 5	434	30.1	434	9	AI991272	AI991272 wu41h04.x
C 6	414	28.7	414	9	AA593860	AA593860 ml19f03.s
C 7	405	28.1	405	9	AI274929	AI274929 q149c11.x
C 8	389	27.0	505	9	AI791844	AI791844 nk08a11.y
C 9	388	26.9	538	9	AI660493	AI660493 we67h02.x
C 10	382	26.5	411	9	AI983793	AI983793 wu20c09.x
C 11	381	26.4	381	9	AI281211	AI281211 qk58e08.x
C 12	375	26.0	573	14	CB854121	CB854121 UI-CF-DU1
C 13	374	26.0	405	9	AI346155	AI346155 qp43f12.x
C 14	369	25.6	486	9	AI660560	AI660560 we68h08.x
C 15	348	24.1	737	14	CB305339	CB305339 UI-CF-EN1
C 16	330	22.9	466	9	AA573825	AA573825 nk08a11.s
C 17	317	22.0	324	12	BM987789	BM987789 UI-H-CO0-
C 18	296	20.5	296	9	AI821178	AI821178 ne17a03.y
C 19	291	20.2	348	9	AM050605	AM050605 w219b11.x
C 20	280	19.4	690	12	BM977010	BM977010 UI-CF-EN1
C 21	277	19.2	277	9	AI732165	AI732165 ne17a03.x
C 22	270	18.7	687	14	CB850544	CB850544 UI-CF-EN1
C 23	260	18.0	260	9	AM009962	AM009962 ws68h06.x
C 24	252	17.5	252	9	AI262416	AI262416 qk38e04.x
C 25	185	12.8	539	9	AI924216	AI924216 wu03b10.x
C 26	183	12.7	593	12	BM973444	BM973444 UI-CF-ECL
C 27	166	11.5	315	9	AA618335	AA618335 ng15g11.s
C 28	149	10.3	274	12	BO017315	BO017315 UI-H-DT1-
C 29	94	6.5	602	9	AA469031	AA469031 ne17a03.s
C 30	88	6.1	151	9	AI695625	AI695625 we50h04.x
C 31	88	6.1	378	10	BF917041	BF917041 IL3-VT011
C 32	44	3.1	119	9	AI921289	AI921289 wo22h06.x
C 33	42	2.9	942	14	CF241985	CF241985 AGENCOURT
C 34	41	2.8	279	9	AI270429	AI270429 qu64h03.x
C 35	40	2.8	175	14	CF316906	CF316906 HD--06-G1
C 36	40	2.8	217	14	CF328864	CF328864 NACL--03-
C 37	39	2.7	188	12	BM887036	BM887036 sam33d08.
C 38	39	2.7	225	9	AI932794	AI932794 wo36g03.x
C 39	39	2.7	243	13	BQ394290	BQ394290 NISC-ng08
C 40	39	2.7	248	9	AI537827	AI537827 lp35h10.x
C 41	39	2.7	250	14	CD421839	CD421839 laa66d09.
C 42	39	2.7	284	9	AI537819	AI537819 lp35g10.x
C 43	39	2.7	300	9	AI475331	AI475331 cl81h09.x
C 44	39	2.7	324	10	AM130187	AM130187 xf29f09.x
C 45	39	2.7	328	13	BU055639	BU055639 UI-M-F00-

ALIGNMENTS

RESULT 1
AI821606/c 537 bp mRNA linear EST 13-DEC-1999
nk08a11.x5 NCI CGAP Co2 Homo sapiens CDNA clone IMAGE:1012892 3'
similar to contains_TARI.t3 TARI MERR2 repetitive element ; , mRNA
sequence.

ACCESSION AI821606
VERSION AI821606.1 GI:5440685
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 537)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT Other ESTs: nk08a11.y5
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc., David B. Krizman,
Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: www-bio.lnl.gov/bbrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: National Cancer Institute, Cancer Genome
Anatomy Project (CGAP), Tumor Gene Index
This read has been verified (found to hit its original self in the
correct orientation)
Insert Length: 741 Std Error: 0.00
Seq primer: -40UP from G1bco
High quality sequence stop: 459.

FEATURES

source

```
1.537
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1012892"
/tissue_type="tumor"
/lab_host="SOLR (Kanamycin resistant)"
/clone_lib="NCI CGAP_CO2"
/note="Organ: colon; Vector: Bluescript SK-; Site: 1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Bulk colon villosus adenoma. 5' adaptor sequence:
5' GAATTCGGCAGCAG 3' 3' adaptor sequence: 5'
CTCGATTCTTTTCTTTTCTTTT 3' Average insert size: 1.1 Kb."
```

ORIGIN

Query Match

Best Local Similarity 36.6%; Score 528; DB 9; Length 537;
Matches 528; Conservativity 100.0%; Pred. No. 1.6e-95;
Mismatch 0; Indels 0; Gaps 0;

```
QY 878 GCCTGGCACTGCTGACCAACCGAGCCTTCCGCTCTTCGGGCTTCGCTTGGCCTCCA 937
Db 537 GCCTGGCACTGCTGACCAACCGAGCCTTCCGCTCTTCGGGCTTCGCTTGGCCTCCA 478
QY 938 TCTCTAGCGTGCCTCTGCGCTCCGCTAGGCTCCTCCGCGCTCACCACTCAGTACG 997
Db 477 TCTCTAGCGTGCCTCTGCGCTCCGCTAGGCTCCTCCGCGCTCACCACTCAGTACG 418
QY 998 GCGCGCCTTCTGCGTCAAGCTGGCAACCGGCTCTGCTCTTCTCTCGGAGGGCGCG 1057
Db 417 GCGCGCCTTCTGCGTCAAGCTGGCAACCGGCTCTGCTCTTCTCTCGGAGGGCGCG 358
QY 1058 TGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTCGACACCTTCTTGACCAAGCGCCA 1117
Db 357 TGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTCGACACCTTCTTGACCAAGCGCCA 298
QY 1118 AGGACTGCAGCCAGAGAGAGGGGCTCACCTCTTATCTCTCGGAGCCCACTGCACAGC 1177
Db 297 AGGACTGCAGCCAGAGAGAGGGGCTCACCTCTTATCTCTCGGAGCCCACTGCACAGC 238
QY 1178 AGGCGCTCTCCAGACTTAAATGATACCACTAACCCTGTGAGGGGAGCCCAATCTGG 1237
Db 237 AGGCGCTCTCCAGACTTAAATGATACCACTAACCCTGTGAGGGGAGCCCAATCTGG 178
QY 1238 ACTCTTCCCCCGCTTGGGACATCGCAGGCCGCGGAGCAGTGCCTCCAGGCTGGGCCA 1297
Db 177 ACTCTTCCCCCGCTTGGGACATCGCAGGCCGCGGAGCAGTGCCTCCAGGCTGGGCCA 118
QY 1298 GGAGAGCTCCAGGAAGGGCACTGAGCGCTGCTGGCGGAGGCGCTTCGACATCCGAGGCA 1357
Db 117 GGAGAGCTCCAGGAAGGGCACTGAGCGCTGCTGGCGGAGGCGCTTCGACATCCGAGGCA 58
QY 1358 CCAGGGAAGTCTCTCTGGGGCGATCTGTAATAAATCTTTTCTTT 1405
Db 57 CCAGGGAAGTCTCTCTGGGGCGATCTGTAATAAATCTTTTCTTT 10
```

RESULT 2
A1826629/c A1826629 515 bp mRNA linear EST 21-DEC-1999
LOCUS
DEFINITION wk35e04.x1 NCI CGAP_Pr22 Homo sapiens cDNA clone IMAGE:2417406 3',

ACCESSION A1826629
VERSION A1826629.1 GI:5447300
KEYWORDS EST.

SOURCE

ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 515)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps@mail.nih.gov

Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:

www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 910 Std Error: 0.00

Seq primer: -40UP from G1bco
High quality sequence stop: 467.

FEATURES

source

```
1.515
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2417406"
/sex="male"
/tissue_type="normal prostate"
/lab_host="DH10B"
/clone_lib="NCI CGAP_Pr22"
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; 1st strand cDNA was prepared
from normal prostate bulk tissue, and was then primed with
a Not I - oligo(dt) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library is normalized, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

ORIGIN

Query Match

Best Local Similarity 35.7%; Score 515; DB 9; Length 515;
Matches 515; Conservativity 100.0%; Pred. No. 6.3e-93;
Mismatch 0; Indels 0; Gaps 0;

```
QY 902 CTTGCGGCTCTTGGGGCTTCTGCGCTTGGCCTTCACTCTAGCGTCCGCTTGGCCGC 961
Db 515 CTTGCGGCTCTTGGGGCTTCTGCGCTTGGCCTTCACTCTAGCGTCCGCTTGGCCGC 456
QY 962 TCCGCTTAGGCTCTCCGGGCTCACCACTCAGTACGGCGCCGCTTCTGGTACGCTGG 1021
Db 455 TCCGCTTAGGCTCTCCGGGCTCACCACTCAGTACGGCGCCGCTTCTGGTACGCTGG 396
QY 1022 CAACCGGCTCTGTGCTCTTCTCTCGGAGGGGCGGCTGTGAGTCTCCAGTATGTTCCGC 1081
Db 455 TCCGCTTAGGCTCTCCGGGCTCACCACTCAGTACGGCGCCGCTTCTGGTACGCTGG 396
QY 395 CAACCGGCTCTGTGCTCTTCTCTCGGAGGGGCGGCTGTGAGTCTCCAGTATGTTCCGC 336
Db 1082 CCAGCGCTCTTCCGACCCCTTCTTGACCAAGCGCCAGGACTGCAGCCAGAGAGGGG 1141
QY 335 CCAGCGCTCTTCCGACCCCTTCTTGACCAAGCGCCAGGACTGCAGCCAGAGAGGGG 276
Db 1142 GCTCACCTCTTATCTCTGGGAGCCCACTGCACAAGCGCGCTTCCAGACTTAAAT 1201
QY 275 GCTCACCTCTTATCTCTGGGAGCCCACTGCACAAGCGCGCTTCCAGACTTAAAT 216
Db 1202 GTATCACCACTAAGTGTGAGGGGAGCAACCAATCTGAGTCTTCCCGCTTGGGACATC 1261
QY 215 GTATCACCACTAAGTGTGAGGGGAGCAACCAATCTGAGTCTTCCCGCTTGGGACATC 156
```

QY 1262 GCAGCGCGGAGACAGTCCCGCCAGGCTGGCCAGAGAGCTCCAGGAAGGCACCTGA 1321
 |||
 DB 155 GCAGCGCGGAGAGCAGTGGCCCGCCAGGCTGTGGCCAGAGAGCTCCAGGAAGGCACCTGA 96
 |||
 QY 1322 GCGCTGTGGCGCGAGGCTCGACATCCGAGCGACCAAGAACTCTCTGGGGCGAT 1381
 |||
 DB 95 GCGCTGTGGCGCGAGGCTCGACATCCGAGCGACCAAGAACTCTCTGGGGCGAT 36
 |||
 QY 1382 CTGTAATAAACCTTTTCTTTCTTTGTTTAA 1416
 |||
 DB 35 CTGTAATAAACCTTTTCTTTCTTTGTTTAA 1

RESULT 3

AI304327/c 510 bp mRNA linear EST 01-FEB-1999
 LOCUS q057h06.x1 NCI_CGAP Co8 Homo sapiens cDNA clone IMAGE:1912667 3'
 DEFINITION similar to contains TAR1.t3 TAR1 repetitive element ;, mRNA
 sequence.

ACCESSION AI304327
 VERSION AI304327.1 GI:3988016
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 510)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bdrp/image/image.html
 Insert Length: 1487 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 476.
 Location/Qualifiers

FEATURES

source

1..510
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1912667"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Co8"
 /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 colon adenocarcinoma, and was then primed with a Not I -
 oligo(dt) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT7T3
 vector. Library is normalized. Library was constructed by
 Bento Soares and M. Fatima Bernaldo."

ORIGIN

Query Match 35.4%; Score 510; DB 9; Length 510;
 Best Local Similarity 100.0%; Pred. No. 6.2e-92;
 Matches 510; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 904 TTGCGGCTTTGGGGGTCTTGGCCTTGCATCTCTAGCGCGCTGTGCGGCTC 963
 |||
 DB 510 TTGCGGCTTTGGGGGTCTTGGCCTTGCATCTCTAGCGCGCTGTGCGGCTC 451
 |||
 QY 964 CGCCTAGGCTCTCCGCGCTCACCACTCAGTAGCGCGCGCTTCTGGGTACGCGCTGCA 1023

DB 450 CGCCTAGGCTCTCCGCGCTCACCACTCAGTAGCGCGCGCTTCTGGGTACGCGTGGCA 391
 |||
 QY 1024 ACCGGGCTCTGTGCTCTTCTCTCGAGAGGGCCGTGGTAGTCTCCAGTATGTTGGGCC 1083
 |||
 DB 390 ACCGGGCTCTGTGCTCTTCTCTCGAGAGGGCCGTGGTAGTCTCCAGTATGTTGGGCC 331
 |||
 QY 1084 AGCGCTCTTGGCACTCTTGTGACCAAGCCCAAGGACTGACGCCAGAGAGAGAGGGGCC 1143
 |||
 DB 330 AGCGCTCTTGGCACTCTTGTGACCAAGCCCAAGGACTGACGCCAGAGAGAGAGGGGCC 271
 |||
 QY 1144 TCACCTCTTATCTCTGCGGCACTGACCAAGAGGCGCTCTCCAGACTTAAATGT 1203
 |||
 DB 270 TCACCTCTTATCTCTGCGGCACTGACCAAGAGGCGCTCTCCAGACTTAAATGT 211
 |||
 QY 1204 ATCACCACCTAATCTGTGAGGGGGACCAATCTGACTCTTCCCGCTTGGAGATCGC 1263
 |||
 DB 210 ATCACCACCTAATCTGTGAGGGGGACCAATCTGACTCTTCCCGCTTGGAGATCGC 151
 |||
 QY 1264 AGGCGGGAGAGAGTGCCTGCGGCAAGGCTGGGCGAGAGCTCCAGAAAGGGCACTGAGC 1323
 |||
 DB 150 AGGCGGGAGAGAGTGCCTGCGGCAAGGCTGGGCGAGAGCTCCAGAAAGGGCACTGAGC 91
 |||
 QY 1324 GCTGCTGGCGGAGGCTTGGACATCCGAGGCAAGGAAAGTCTCTGGGGCGATCT 1383
 |||
 DB 90 GCTGCTGGCGGAGGCTTGGACATCCGAGGCAAGGAAAGTCTCTGGGGCGATCT 31
 |||
 QY 1384 GTAATAAACCTTTTCTTTCTTTGTTT 1413
 |||
 DB 30 GTAATAAACCTTTTCTTTCTTTGTTT 1

RESULT 4

AI984141/c 446 bp mRNA linear EST 27-OCT-1999
 LOCUS wu21c02.x1 Soares_Dieckgraefe_colon_NHCD Homo sapiens cDNA clone
 DEFINITION IMAGE:2520674 3', mRNA sequence.

ACCESSION AI984141
 VERSION AI984141.1 GI:5811360
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 446)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40UP from Gibco.

Location/Qualifiers
 1..446
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2520674"
 /tissue_type="colonic mucosa from 3 patients with Crohn's
 disease"

FEATURES

source

/lab_host="DH10B (phage-resistant)"
 /clone_lib="Soares_Dieckgraefe_colon_NHCD"
 /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
 modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dt) primer [5'
 TGTTACCAATCTGAAAGTGGAGCGCGCGCTTTTCTTTTCTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization. Tissue samples
 provided by Dr. Brian Dieckgraefe (Washington University,

dieck@m.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 31.0%; Score 446; DB 9; Length 446;
Best Local Similarity 100.0%; Pred. No. 3e-79;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 971 GCTCTCCGGCTCACCACCTAGTACGGCGCCGCTTCTGGGTACGCTGGCAACCGCG 1030
DB 446 GCTCTCCGGCTCACCACCTAGTACGGCGCCGCTTCTGGGTACGCTGGCAACCGCG 387
QY 1031 TCCTGTGCTCTCTCCGAGGGGCGGTGTAGTCTCCAGTATGTTCCGCCAGCGCTC 1090
DB 386 TCCTGTGCTCTCTCCGAGGGGCGGTGTAGTCTCCAGTATGTTCCGCCAGCGCTC 327
QY 1091 TTCCGACCCCTTCTGACCAAGCGCCAGACTGACCCAGAGAGAGGGGCTCACCTC 1150
DB 326 TTCCGACCCCTTCTGACCAAGCGCCAGACTGACCCAGAGAGAGGGGCTCACCTC 267
QY 1151 TTATCTCTCGGCGACCCACCTGACCAAGAGGCGCTCTCCAGACTTAAATGTATCACC 1210
DB 266 TTATCTCTCGGCGACCCACCTGACCAAGAGGCGCTCTCCAGACTTAAATGTATCACC 207
QY 1211 CTAACCTGTAGAGGGGAGCCCAATCTGACTCTTCCCGCCCTTGGGACATCGAGGCGCG 1270
DB 206 CTAACCTGTAGAGGGGAGCCCAATCTGACTCTTCCCGCCCTTGGGACATCGAGGCGCG 147
QY 1271 GAAGCAGTGGCCCGCCAGGCTGGGCCAGAGAGAGCTCCAGAGAGGGGCACTGAGCGCTG 1330
DB 146 GAAGCAGTGGCCCGCCAGGCTGGGCCAGAGAGAGCTCCAGAGAGGGGCACTGAGCGCTG 87
QY 1331 GCGCGAGGCTCGGACATCGGACAGGACAGGAAAGTCTCCTGGGGCGATCTGTAATA 1390
DB 86 GCGCGAGGCTCGGACATCGGACAGGACAGGAAAGTCTCCTGGGGCGATCTGTAATA 27
QY 1391 AACCTTTTCTTTCTTTTAA 1416
DB 26 AACCTTTTCTTTCTTTTAA 1

RESULT 5
AI991272/c 434 bp mRNA linear EST 09-MAR-2000
LOCUS wu41h04.x1 Soares_Dieckgraeft_colon_NHCD Homo sapiens cDNA clone
DEFINITION IMAGE:2522647 3', mRNA sequence.

ACCESSION AI991272
VERSION AI991272.1 GI:5838177
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 434)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 558 Std Error: 0.00
Seq primer: -40UP from Gibco.

FEATURES

source
1. 434
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2522647"

/tissue_type="colonic mucosa from 3 patients with Crohn's disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares_Dieckgraeft_colon_NHCD"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraeft (Washington University, dieck@m.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 30.1%; Score 434; DB 9; Length 434;
Best Local Similarity 100.0%; Pred. No. 7.2e-77;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 981 GCTCACCACCTAGTACGGCGCCGCTTCTGGGTACGCTGGCAACCGCGCTCTGCTGCT 1040
DB 434 GCTCACCACCTAGTACGGCGCCGCTTCTGGGTACGCTGGCAACCGCGCTCTGCTGCT 375
QY 1041 CTTCTCTGAGAGGGGCGGTGTAGTCTCAGTATGTTGGCCAGCGCTCTTGGACCT 1100
DB 374 CTTCTCTGAGAGGGGCGGTGTAGTCTCAGTATGTTGGCCAGCGCTCTTGGACCT 315
QY 1101 TCTGACCAAAAGCGCCAGGACTGACGACAGAGAGAGGGGCTCACTTATCTCTGG 1160
DB 314 TCTGACCAAAAGCGCCAGGACTGACGACAGAGAGAGGGGCTCACTTATCTCTGG 255
QY 1161 CGAACCACCTGACCAAGAGCGGCTCTCCAGACTTAAATGTATACCACTAAGCTGTG 1220
DB 254 CGAACCACCTGACCAAGAGCGGCTCTCCAGACTTAAATGTATACCACTAAGCTGTG 195
QY 1221 AGGGGGAGCCCAATCTGACTCTTCCCGGCTTGGGACATCGCAGGCGGGAGAGCTGC 1280
DB 194 AGGGGGAGCCCAATCTGACTCTTCCCGGCTTGGGACATCGCAGGCGGGAGAGCTGC 135
QY 1281 CCGCCAGGCTTGGGCCAGAGAGACTCCAGAAAGGCACTGAGCGCTGTGGCGGAGGCC 1340
DB 134 CCGCCAGGCTTGGGCCAGAGAGACTCCAGAAAGGCACTGAGCGCTGTGGCGGAGGCC 75
QY 1341 TCGGACATCCGACAGGACCAAGGAAGTCTCCTGGGGGATCTGTAATAAACCTTTT 1400
DB 74 TCGGACATCCGACAGGACCAAGGAAGTCTCCTGGGGGATCTGTAATAAACCTTTT 15
QY 1401 TCTTTTGTTTT 1414
DB 14 TCTTTTGTTTT 1

RESULT 6
AA593860/c 414 bp mRNA linear EST 25-SEP-1997
LOCUS un19f03.s1 NCI_CGAP_C012 Homo sapiens cDNA clone IMAGE:1084349 3',
DEFINITION mRNA sequence.

ACCESSION AA593860
VERSION AA593860.1 GI:2408538
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 414)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:
www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 1204 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 406.

FEATURES

source
1. .414
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1084349"
/sex="mixed"
/tissue_type="colon tumor"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="NCI_CGAP_Col2"
/note="Organ: colon; Vector: Bluescript SK-; Site 1: EcCR1; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Pooled colon tumors. 5' adaptor sequence: 5' GAATTCGGCAGCAG 3', 3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3', Average insert size: 1.2 kb."

ORIGIN

Query Match 28.7%; Score 414; DB 9; Length 414;
Best Local Similarity 100.0%; Pred. No. 6.7e-73;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 999 CGCCGCTTCTGGGTACGCTGGCAACCGCGCTCTGCTCTTCTCGAGGGCGCT 1058
DB 414 CGCCGCTTCTGGGTACGCTGGCAACCGCGCTCTGCTCTTCTCGAGGGCGCT 355
QY 1059 GGTGAGTCTCCAGTATGTTCCGCCCAAGCGCTTTTGCACCTTCTGACCAAGCGCAA 1118
DB 354 GGTGAGTCTCCAGTATGTTCCGCCCAAGCGCTTTGACACCTTCTGACCAAGCGCAA 295
QY 1119 GGAATGACGACGAGAGAGAGGGGGCTCACTCTTATCTCTGGCGACCACTGACAGCA 1178
DB 294 GGAATGACGACGAGAGAGAGGGGGCTCACTCTTATCTCTGGCGACCACTGACAGCA 235
QY 1179 GGGCGCTCTCCAGACTTAATATGATACCACTAAGCTGTGAGGGGAGCCCAATCTGA 1238
DB 234 GGGCGCTCTCCAGACTTAATATGATACCACTAAGCTGTGAGGGGAGCCCAATCTGA 175
QY 1239 CTCCTTCCCGCTTGGACATCGCAGCGCGGAGCAAGTACCGCCGAGGCTGGGCCAG 1298
DB 174 CTCCTTCCCGCTTGGACATCGCAGCGCGGAGCAAGTACCGCCGAGGCTGGGCCAG 115
QY 1299 GAGAGCTCCAGAGAGGGGCACTGAGCGCTGTGCGCGGAGGCTCGACATCCGACAGCAC 1358
DB 114 GAGAGCTCCAGAGAGGGGCACTGAGCGCTGTGCGCGGAGGCTCGACATCCGACAGCAC 55
QY 1359 CAGGAGAAAGTCTCTCTGGGGCGATCTGTAATAAACCTTTTCTTTGTTT 1412
DB 54 CAGGAGAAAGTCTCTCTGGGGCGATCTGTAATAAACCTTTTCTTTGTTT 1

RESULT 7

AI274929/c 405 bp mRNA linear EST 29-JAN-1999
LOCUS AI274929
DEFINITION q149c11.x1 NCI_CGAP_C08 Homo sapiens cDNA clone IMAGE:1875668 3'
similar to contains TAR1.t3 TAR1 repetitive element ;, mRNA
sequence.
ACCESSION AI274929
VERSION AI274929.1 GI:3897203
KEYWORDS EST.

SOURCE ORGANISM

Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 405)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)

JOURNAL COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:
www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 1458 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 395.

FEATURES

source
1. .405
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1875668"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_C08"
/note="Organ: colon; Vector: pT7UD-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bernaldo."

ORIGIN

Query Match 28.1%; Score 405; DB 9; Length 405;
Best Local Similarity 100.0%; Pred. No. 4.1e-71;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1009 TGGGTACGCTGGCAACCGCGCTCTGCTCTTCTCGAGGGCGCTGAGTCTC 1068
DB 405 TGGGTACGCTGGCAACCGCGCTCTGCTCTTCTCGAGGGCGCTGAGTCTC 346
QY 1069 CAGTATGTTCCGCCAGCGCTCTTGCACCGCTTCTGACCAAGCGCCAGACTGCAGC 1128
DB 345 CAGTATGTTCCGCCAGCGCTCTTGCACCGCTTCTGACCAAGCGCCAGACTGCAGC 286
QY 1129 CAGGAGAGGGGGCTCACTCTTATCTCTGGGAGCCCACTGACAGAGGCGCTCTC 1188
DB 285 CAGGAGAGGGGGCTCACTCTTATCTCTGGGAGCCCACTGACAGAGGCGCTCTC 226
QY 1189 CCAGACTTAATATGATACCACTAAGCTGTGAGGGGAGCCCAATCTGACTCTTCCC 1248
DB 225 CCAGACTTAATATGATACCACTAAGCTGTGAGGGGAGCCCAATCTGACTCTTCCC 166
QY 1249 GCCTTGGACATCGCAGGCGCGGAGAGCAAGTACCGCCGAGGCTGGGCCAGAGAGTCCA 1308
DB 165 GCCTTGGACATCGCAGGCGCGGAGAGCAAGTACCGCCGAGGCTGGGCCAGAGAGTCCA 106
QY 1309 GGAAGGCACTGAGCGCTGTGGGGGAGGCGCTTGGACATCCGACAGGCAAGGAAAGT 1368
DB 105 GGAAGGCACTGAGCGCTGTGGGGGAGGCGCTTGGACATCCGACAGGCAAGGAAAGT 46
QY 1369 CTCCTGGGGCGATCTGTAATAAACCTTTTCTTTGTTT 1413
DB 45 CTCCTGGGGCGATCTGTAATAAACCTTTTCTTTGTTT 1

RESULT 8
LOCUS AI791844 505 bp mRNA linear EST 13-DEC-1999
DEFINITION nk08a11.y5 NCI CGAP Co2 Homo sapiens CDNA clone IMAGE:1012892 5' similar to contains_Alu repetitive element; mRNA sequence.
ACCESSION AI791844
VERSION AI791844
KEYWORDS AI791844.1 GI:5339486
SOURCE EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 505)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Other_ESTS: nk08a11.s1
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc., David B. Krizman, Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/dbp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: see original entry for original citation
information
This 5' resequenced clone has no previous 5' data to verify this
new read against
Putative full length read
The vector to vector length is 671
Insert Length: 741 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 418.
Location/Qualifiers
1..505
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1012892"
/tissue_type="tumor"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="NCI CGAP Co2"
/note="Organ: colon; Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI; cloned unidirectionally. Primer: Oligo dt. Bulk colon villous adenoma. 5' adaptor sequence: 5' GAATTCGGCAGAG 3' 3' adaptor sequence: 5' CTCGAGTTT TTT TTT TTT TTT 3' Average insert size: 1.1 kb."

ORIGIN
Query Match 27.0%; Score 389; DB 9; Length 505;
Best Local Similarity 100.0%; Pred. No. 4.8e-68;
Matches 389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 736 AAGTTCACACCGAGTAGACCCCTGCGGCTGTACACACAGTACACCTGGCGGACACTAC 795
Db 1 AAGTTCACACCGAGTAGACCCCTGCGGCTGTACACACAGTACACCTGGCGGACACTAC 60

QY 796 GCGTCGGCCACGCTATGGGTGGGCTTCTGCTTCTGCTCTCTCCAAAGTGTGCTCTCC 855
Db 61 GCGTCGGCCACGCTATGGGTGGGCTTCTGCTTCTGCTCTCTCCAAAGTGTGCTCTCC 120

QY 856 ACGCGGGCCCCGCTCTACGAGGCGCTGGACCTGACCAACGGAGCCTTCGGCTCTTC 915
Db 121 ACGCGGGCCCCGCTCTACGAGGCGCTGGACCTGACCAACGGAGCCTTCGGCTCTTC 180

QY 916 GGGGTCTTGGCCTTGGCCTCCATCTAGCGTGGCCGCTCGCCCGCTCCGCTAGAGCTCC 975
Db 181 GGGGTCTTGGCCTTGGCCTCCATCTAGCGTGGCCGCTCGCCCGCTCCGCTAGAGCTCC 240

QY 976 TCCGCGCTCACCACCTAGTACGCGCGCCGCTTGGGTACGCTGGCAACCGGCTCTG 1035
Db 241 TCCGCGCTCACCACCTAGTACGCGCGCCGCTTGGGTACGCTGGCAACCGGCTCTG 300

QY 1036 TGCCTCTTCTCGGAGGGCGCGTGGTGTCTCCAGTATGTCGCCACGCTCTTCCG 1095
Db 301 TGCCTCTTCTCGGAGGGCGCGTGGTGTCTCCAGTATGTCGCCACGCTCTTCCG 360

QY 1096 ACCCTTCTGACCAAGCGCCCAAGGACTG 1124
Db 361 ACCCTTCTGACCAAGCGCCCAAGGACTG 389

RESULT 9
LOCUS AI660493/c 538 bp mRNA linear EST 18-DEC-1999
DEFINITION we67h02.x1 Soares Dieckgraefe colon NHCD Homo sapiens CDNA clone IMAGE:2346195 3' similar to contains TAR1.t3 TAR1 repetitive element; mRNA sequence.
ACCESSION AI660493 GI:4764063
VERSION AI660493.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 538)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 676 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 463.
Location/Qualifiers
1..538
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2346195"
/tissue_type="colonic mucosa from 3 patients with Crohn's disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares_Dieckgraefe_colon NHCD"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5' GTTACCAATCTGAGTGGAGCGCGCGCTTTT TTT TTT TTT TTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieck@lm.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 26.9%; Score 388; DB 9; Length 538;
Best Local Similarity 100.0%; Pred. No. 7.1e-68;
Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1027 GCGTCTGTGCTTCTCTCGAGGGCGCGTGTGAGTCTCCAGTATGTCGCCAGC 1086

```

Db      388  GGGCTCTGTGCTCTTCTTCTCGAGAGGCGGTGTGAGTCTCCAGTATGTTCCGCCAGC 329
QY      1087  GCTCTTCGACCCCTTCTTGACCCAAAGCCCAAGACTGCAGCCAGAGAGAGGGGGCTCA 1146
Db      328  GCTCTTCGACCCCTTCTTGACCCAAAGCCCAAGACTGCAGCCAGAGAGAGGGGGCTCA 269
QY      1147  CCTCTTATCTCGGCGACCACTGCACAAGCAGCGCTCTCCACACTTAATATGATC 1206
Db      268  CCTCTTATCTCGGCGACCACTGCACAAGCAGCGCTCTCCACACTTAATATGATC 209
QY      1207  ACCACTAACCCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGGACATCGCAGG 1266
Db      208  ACCACTAACCCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGGACATCGCAGG 149
QY      1267  CCGGGAAGCAGTGGCCCGCCAGGCTGGGCGAGAGAGCTCCAGAAAGGCACTGAGCGCT 1326
Db      148  CCGGGAAGCAGTGGCCCGCCAGGCTGGGCGAGAGAGCTCCAGAAAGGCACTGAGCGCT 89
QY      1327  GCTGGCGGAGGCTCGGACATCCGACAGGCAACAGGAAAGTCTCTGGGGCGATCTGTA 1386
Db      88  GCTGGCGGAGGCTCGGACATCCGACAGGCAACAGGAAAGTCTCTGGGGCGATCTGTA 29
QY      1387  AATAAACCTTTTCTTTCTTTGTTTGA 1414
Db      28  AATAAACCTTTTCTTTCTTTGTTTGA 1

```

```

RESULT 10
AI983793/c 411 bp mRNA linear EST 27-OCT-1999
LOCUS      AI983793
DEFINITION w120c09.x1 Soares Dieckgraefe colon_NHCD Homo sapiens cDNA clone
ACCESSION  AI983793
VERSION     AI983793
KEYWORDS   AI983793.1 GI:5811012
SOURCE     EST.
ORGANISM   Homo sapiens (human)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 411)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LNL; contact the
            IMAGE Consortium (info@image.lnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 400.
FEATURES   Location/Qualifiers
            source          1..411
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:2520592"
                        /tissue_type="colonic mucosa from 3 patients with Crohn's
                        disease"
                        /lab_host="DH10B (phage-resistant)"
                        /clone_11b="Soares_Dieckgraefe_colon_NHCD"
                        /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
                        modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
                        strand cDNA was primed with a Not I - oligo(dT) primer [5'
                        TGTTACCAATCTGAAGTGGAGCGCGCGCTTTTCTTTTCTTTT 3'],
                        double-stranded cDNA was ligated to Eco RI adaptors
                        (Pharmacia), digested with Not I and cloned into the Not I
                        and Eco RI sites of the modified pT73 vector. Library
                        went through one round of normalization. Tissue samples
                        provided by Dr. Brian Dieckgraefe (Washington University,
                        dieck@im.wustl.edu); colonic mucosa represents a range of
                        disease involvement from moderate to severe Crohn's
                        disease; samples include both perforating (fistulas) and

```

non-perforating samples. Library constructed by Bento Soares and M. Fatima Ronaldo. "

Query Match 26.5%; Score 382; DB 9; Length 411;
 Best Local Similarity 100.0%; Pred. No. 1.4e-66;
 Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1008  CTGGGTACCGCTGGCAACCGGCGTCTGTGCTCTTCTCTCGAGAGGGCGGTGAGTCT 1067
Db      411  CTGGGTACCGCTGGCAACCGGCGTCTGTGCTCTTCTCTCGAGAGGGCGGTGAGTCT 352
QY      1068  CCACTATGTTGGGCCAGCGCTCTTGGACACCTTCTGACCAAGCCCAAGACTGCAG 1127
Db      351  CCACTATGTTGGGCCAGCGCTCTTGGACACCTTCTGACCAAGCCCAAGACTGCAG 292
QY      1128  CCAGAGAGAGGGGCTCACTCTTATCTCTGGCGACCACTGCACAAGAGCGGCTCT 1187
Db      291  CCAGAGAGAGGGGCTCACTCTTATCTCTGGCGACCACTGCACAAGAGCGGCTCT 232
QY      1188  CCCGACTTAATATGTATACCACTAACCCTGTGAGGGGAGCCCAATCTGACTCTTCCC 1247
Db      231  CCCGACTTAATATGTATACCACTAACCCTGTGAGGGGAGCCCAATCTGACTCTTCCC 172
QY      1248  CGCCTTGGACATCGCAGCGCGGGAAGCAGTGCCTCCAGGCTTGGCCAGAGAGCTCC 1307
Db      171  CGCCTTGGACATCGCAGCGCGGGAAGCAGTGCCTCCAGGCTTGGCCAGAGAGCTCC 112
QY      1308  AGGAAGGGCAGCTGAGCGCTGTGGCGGAGGCTCGACATCCGACGCAACAGGAAAG 1367
Db      111  AGGAAGGGCAGCTGAGCGCTGTGGCGGAGGCTCGACATCCGACGCAACAGGAAAG 52
QY      1368  TCTCCTGGGCGGATCTGTAAT 1389
Db      51  TCTCCTGGGCGGATCTGTAAT 30

```

```

RESULT 11
AI281211/c 381 bp mRNA linear EST 28-JAN-1999
LOCUS      AI281211/c
DEFINITION qk58e08.x1 NCI CGAP C08 Homo sapiens cDNA clone IMAGE:1873190 3'
            similar to contains TAR1.t3 TAR1 repetitive element ;, mRNA
            sequence.
ACCESSION  AI281211
VERSION     AI281211.1 GI:3919444
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 381)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            found through the I.M.A.-G.B. Consortium/LNL at:
            www-bio.lnl.gov/bbrp/image/image.html
            Insert Length: 1422 Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 340.
FEATURES   Location/Qualifiers
            source          1..381
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"

```

/clone="IMAGE:1873190"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI CGAP_C08"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 26.4%; Score 381; DB 9; Length 381;
Best Local Similarity 100.0%; Pred. No. 2.4e-66;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1033 CTGTGCTCTTCTCTCGAGAGGGCCGTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTT 1092
DB 381 CTGTGCTCTTCTCTCGAGAGGGCCGTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTT 322
QY 1093 CGCACCTCTTGTGACCAAGCCGCAAGAGTGTGAGCCAGAGAGAGGGGCTCACCTCTT 1152
DB 321 CGCACCTCTTGTGACCAAGCCGCAAGAGTGTGAGCCAGAGAGAGGGGCTCACCTCTT 262
QY 1153 ATCTCTCGGCGACCCACTGACACAGAGAGGCGCTCTCCAGACTTAAATGTATCCAGCACT 1212
DB 261 ATCTCTCGGCGACCCACTGACACAGAGAGGCGCTCTCCAGACTTAAATGTATCCAGCACT 202
QY 1213 AACCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCCTTGGGACATCGAGGCGCGGA 1272
DB 201 AACCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCCTTGGGACATCGAGGCGCGGA 142
QY 1273 AGCAGTGCCTCCGCGAGGCTGGCCAGAGAGTCCAGAGAGGCACTGAGCGCTGTGCG 1332
DB 141 AGCAGTGCCTCCGCGAGGCTGGCCAGAGAGTCCAGAGAGGCACTGAGCGCTGTGCG 82
QY 1333 GCGAGGCTCTCGACATCCGCGAGGAGGAAAGTCTCTGGGCGATCTGTAATATA 1392
DB 81 GCGAGGCTCTCGACATCCGCGAGGAGGAAAGTCTCTGGGCGATCTGTAATATA 22
QY 1393 CCTTTTCTTTTGTGTTT 1413
DB 21 CCTTTTCTTTTGTGTTT 1

RESULT 12

CB854121/c 573 bp mRNA linear EST 22-APR-2003
LOCUS CB854121
DEFINITION UI-CF-DUI-aal-1-16-0-UI.s1 UI-CF-DUI Homo sapiens cDNA clone
KEYWORDS UI-CF-DUI-aal-1-16-0-UI 3', mRNA sequence.
ACCESSION CB854121 GI:30044498
VERSION CB854121
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
JOURNAL discovery
MEDLINE Genome Res. 6 (9), 791-806 (1996)
PUBMED 97044477
COMMENT 8889548
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
Sequence: 533-573, >Alu (matched complement)
Seq primer: M13 FORWARD
POLYA=No.

FEATURES

source

location/Qualifiers
1..573
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-DUI-aal-1-16-0-UI"
/tissue_type="Primary Lung Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-DUI"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-CF-DUI is a normalized cDNA library containing the following tissue(s): Primary Lung Epithelial Cells The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GGCTGTAGGC.
TAG_SEQ=None found"

ORIGIN

Query Match 26.0%; Score 375; DB 14; Length 573;
Best Local Similarity 100.0%; Pred. No. 2.4e-65;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1044 CCTCGAGGGGCGTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTGCACCTTCT 1103
DB 482 CCTCGAGGGGCGTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTGCACCTTCT 423
QY 1104 GGAACCAAGCGCCAGAGTGTGAGCCAGAGAGAGGGGCTCACTTATCTCGCGGA 1163
DB 422 GGAACCAAGCGCCAGAGTGTGAGCCAGAGAGAGGGGCTCACTTATCTCGCGGA 363
QY 1164 CCCACTGCACAGCAGGCGGCTCTCCAGACTTAAATGTATCACCACTAAGTGTAGG 1223
DB 362 CCCACTGCACAGCAGGCGGCTCTCCAGACTTAAATGTATCACCACTAAGTGTAGG 303
QY 1224 GGAACCAATCTGGAATCTTCCCGCGCTTGGGACATCGAGGCGGGAAGCAGTGCCTG 1283
DB 302 GGAACCAATCTGGAATCTTCCCGCGCTTGGGACATCGAGGCGGGAAGCAGTGCCTG 243
QY 1284 CCAAGCCTGGGCGAGAGAGTCCAGAGAGGCACTGAGGCGTGTGGCGGAGGCGCTCG 1343
DB 242 CCAAGCCTGGGCGAGAGAGTCCAGAGAGGCACTGAGGCGTGTGGCGGAGGCGCTCG 183
QY 1344 GACATCCGCGAGGCAAGGAAAGTCTCTGGGCGGATGTGTAATAAATCTTTTCT 1403
DB 182 GACATCCGCGAGGCAAGGAAAGTCTCTGGGCGGATGTGTAATAAATCTTTTCT 123
QY 1404 TTTGTTTTTAAATA 1418
DB 122 TTTGTTTTTAAATA 108

RESULT 13
AI346155/c 405 bp mRNA linear EST 02-FEB-1999
LOCUS AI346155

DEFINITION qp43f12.x1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1925807 3', similar to contains_TARI.t3 TARI repetitive element ;, mRNA

ACCESSION AI346155

VERSION AI346155.1 GI:4083361

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

KEYWORDS Email: cgapbs-r@mail.nih.gov

SOURCE Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

ORGANISM CDNA Library Preparation: M. Bento Soares, Ph.D.

REFERENCE CDNA Library Arrayed by: Greg Lemmon, Ph.D.

AUTHORS DNA Sequencing by: Washington University Genome Sequencing Center

TITLE Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

JOURNAL Insert Length: 1488 Std Error: 0.00

COMMENT Seq primer: -40UP from Gibco

KEYWORDS High quality sequence stop: 375.

SOURCE location/Qualifiers

1. .405

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:1925807"

/tissue_type="adenocarcinoma"

/lab_host="DH10B"

/clone_lib="NCI_CGAP_Co8"

/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 26.0%; Score 374; DB 9; Length 405;

Best Local Similarity 100.0%; Pred. No. 5.4e-65;

Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1009 TGGGTACGCTGGCAACCGGGCTCTGTGCTCTTCTTCGGAGGGGCGGTGAGTCTC 1068

DB 405 TGGGTACGCTGGCAACCGGGCTCTGTGCTCTTCTTCGGAGGGGCGGTGAGTCTC 346

QY 1069 CAGTATGTTGGGCCCAAGCGCTCTTCGACCCCTTCTGACCAAGCGCCCAAGAGACTGCAGC 1128

DB 345 CAGTATGTTGGGCCCAAGCGCTCTTCGACCCCTTCTGACCAAGCGCCCAAGAGACTGCAGC 286

QY 1129 CAGGAGAGAGGGGCTCAGCTCTTATCCTCGGCGACCCACTGCACAGAGAGCGCTCTC 1188

DB 285 CAGGAGAGAGGGGCTCAGCTCTTATCCTCGGCGACCCACTGCACAGAGAGCGCTCTC 226

QY 1189 CCAGACTTAAATGTATCACCACTAACCTTGAGGGGAGCCCAATCTGACTCCTTCCCC 1248

DB 225 CCAGACTTAAATGTATCACCACTAACCTTGAGGGGAGCCCAATCTGACTCCTTCCCC 166

QY 1249 GCCTTGGACATCGCAGCGCGGGAAGCAGTGCCTCCAGGCTTGCGCAGAGAGAGCTCCA 1308

DB 165 GCCTTGGACATCGCAGCGCGGGAAGCAGTGCCTCCAGGCTTGCGCAGAGAGAGCTCCA 106

QY 1309 GGAAGGGCACTGAGCGCTGCTGGCGGAGGCGCTCGACATCCGACAGGCAACGAGGAAGT 1368

Db 105 GGAAGGGCACTGAGCGCTGCTGGCGGAGGCTCGACATCCGACAGGCAACGAGGAAGT 46

QY 1369 CTCCTGGGGCGATC 1382

Db 45 CTCCTGGGGCGATC 32

RESULT 14

LOCUS AI660560/c 486 bp mRNA linear EST 18-DEC-1999

DEFINITION we68b08.x1 Soares Dieckgraefe colon NHCD Homo sapiens cDNA clone IMAGE:2346231 3', similar to contains_TARI.t3 TARI repetitive element ;, mRNA sequence.

ACCESSION AI660560

VERSION AI660560.1 GI:4764130

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

KEYWORDS Email: cgapbs-r@mail.nih.gov

SOURCE This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

ORGANISM Insert Length: 1433 Std Error: 0.00

REFERENCE Seq primer: -40UP from Gibco

AUTHORS High quality sequence stop: 447.

TITLE location/Qualifiers

JOURNAL 1. .486

COMMENT /organism="Homo sapiens"

KEYWORDS /mol_type="mRNA"

SOURCE /db_xref="taxon:9606"

ORGANISM /clone="IMAGE:2346231"

REFERENCE /tissue_type="colonic mucosa from 3 patients with Crohn's disease"

AUTHORS /lab_host="DH10B (phage-resistant)"

TITLE /clone_lib="Soares_Dieckgraefe_colon_NHCD"

JOURNAL /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGGCGGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieck@im.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 25.6%; Score 369; DB 9; Length 486;

Best Local Similarity 99.8%; Pred. No. 4.4e-64;

Matches 419; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1009 TGGGTACGCTGGCAACCGGCGTCTGTGCTCTTCTTCGGAGGGGCGGTGAGTCTC 1068

DB 420 TGGGTACGCTGGCAACCGGCGTCTGTGCTCTTCTTCGGAGGGGCGGTGAGTCTC 361

QY 1069 CAGTATGTTGGGCCCAAGCGCTCTTCGACCCCTTCTGACCAAGCGCCCAAGAGACTGCAGC 1128

DB 360 CAGTATGTTGGGCCCAAGCGCTCTTCGACCCCTTCTGACCAAGCGCCCAAGAGACTGCAGC 301

QY 1129 CAGGAGAGAGGGGCTCAGCTCTTATCCTCGGCGACCCACTGCACAGAGAGCGCTCTC 1188

DB 300 CAGGAGAGAGGGGCTCAGCTCTTATCCTCGGCGACCCACTGCACAGAGAGCGCTCTC 241


```

QY 1189 CCAGACTTAAATGTATCACCACCTAACCCTGTGAGGGGAGACCCCAATCTGACTCTCCCTCC 1248
    |||
Db 240 CCAGACTTAAATGTATCACCACCTAACCCTGTGAGGGGAGACCCCAATCTGACTCTCCCTCC 181
QY 1249 GCCTTGGGACATCGCAGGCGGGAGAGAGTGCCTCCAGGCGCTGGGCGCAGAGAGCTCCA 1308
    |||
Db 180 GCCTTGGGACATCGCAGGCGGGAGAGAGTGCCTCCAGGCGCTGGGCGCAGAGAGAGCTCCA 121
QY 1309 GGAAGGGCACTGAGCGCTGCTGCGCGGAGGCGCTCGGACATCCGACAGGACAGGAAAGT 1368
    |||
Db 120 GGAAGGGCACTGAGCGCTGCTGCGCGGAGGCGCTCGGACATCCGACAGGACAGGAAAGT 61
QY 1369 CTCCTGGGGCGATCTGTAATAAACCCTTTTCTTTTCTTTTAAAAA 1428
    |||
Db 60 CTCCTGGGGCGATCTGTAATAAACCCTTTTCTTTTCTTTTAAAAA 1

```

```

RESULT 15
CB305399 737 bp mRNA linear EST 04-MAR-2003
LOCUS
DEFINITION UI-CF-EN1-aed-m-17-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
CB305399
VERSION CB305399.1 GI:28845910
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 737)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
The following repetitive elements were found in this cDNA
Sequence: 29-138, >MIR#SINE/MIR (matched complement) 195-290, >ALU
223-320, >ALU
Seq primer: M13 FORWARD
POLYA=Yes.

```

FEATURES

```

source
Location/Qualifiers
1..737
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-EN1-aed-m-17-0-UI"
/tissue_type="Primary Lung Cystic Fibrosis Epithelial
Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EN1"
/notes="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker, Site_1: EcoR I; Site_2: Not I;
UI-CF-EN1 is a normalized cDNA library containing the
following tissue(s): Primary Lung Cystic Fibrosis
Epithelial Cells. The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

```

1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGGT. TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS 6hr to LPS 24h TAG LIB=UI-CF-EN1 TAG_SEQ=CTGCTCAGGT"

ORIGIN

```

Query Match 24.1%; Score 348; DB 14; Length 737;
Best Local Similarity 100.0%; Pred. No. 4.1e-60;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1027 GCGCTCTGTGCTCTTCTCTCGAGGGCGGTGTAGTCTCCAGTATGTCGGCCAGC 1086
    |||
Db 353 GCGCTCTGTGCTCTTCTCTCGAGGGCGGTGTAGTCTCCAGTATGTCGGCCAGC 412
QY 1087 GCTCTTGCACCTTCTGTGACCAAGCCCAAGAGTGCAGCCAGAGAGAGGGGCTCA 1146
    |||
Db 413 GCTCTTGCACCTTCTGTGACCAAGCCCAAGAGTGCAGCCAGAGAGAGGGGCTCA 472
QY 1147 CCTCTTATCTCTCGGCGACCCACTGCACAGCAGGCGCTCTCCAGACTTAAATGTATC 1206
    |||
Db 473 CCTCTTATCTCTCGGCGACCCACTGCACAGCAGGCGCTCTCCAGACTTAAATGTATC 532
QY 1207 ACCACTAACCCTGTGAGGGGAGACCAATCTGACTCCTTCCCGCTTGGACATCGCAGG 1266
    |||
Db 533 ACCACTAACCCTGTGAGGGGAGACCAATCTGACTCCTTCCCGCTTGGACATCGCAGG 592
QY 1267 CCGGGAAGCAGTGCCTCCGCGAGGCTGGGCGAGAGAGCTCCAGAAAGGCACTGACGCT 1326
    |||
Db 593 CCGGGAAGCAGTGCCTCCGCGAGGCTGGGCGAGAGAGCTCCAGAAAGGCACTGACGCT 652
QY 1327 GCTGGCGGAGGCTCTGACATCCGACAGCACCAGGAAAGTCTCTG 1374
    |||
Db 653 GCTGGCGGAGGCTCTGACATCCGACAGCACCAGGAAAGTCTCTG 700

```

Search completed: February 24, 2004, 03:21:09
Job time : 3670.04 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

OM nucleic - nucleic search, using sw model
Run on: February 23, 2004, 19:34:53 ; Search time 5685.97 Seconds
(without alignments)
10984.465 Million cell updates/sec
Title: US-09-936-456-1
Perfect score: 1441
Sequence: 1 aaagtaacgctacagacag.....aaaaaaaaaaaaaaaa 1441
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544
Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Match	Query Length	DB	ID	Description
1	1441	100.0	1441	6	AX035346	AX035346 Sequence
2	1029	71.4	1740	9	HSM806241	BX537581 Homo sapi
3	711.4	49.4	1354	10	BC031111	BC031111 Mus muscu
4	626.2	43.5	5851	4	AF547266	AF547266 Sus scrofa
5	465.4	32.3	498	6	AX035348	AX035348 Sequence
6	405.4	28.1	166337	2	AC012255	AC012255 Homo sapi
7	405.4	28.1	181312	9	AC091117	AC091117 Homo sapi
8	403.8	28.0	156534	2	AC009700	AC009700 Homo sapi
9	403.8	28.0	171444	2	AC087790	AC087790 Homo sapi
10	396	27.5	2684	6	AX714511	AX714511 Sequence
11	396	27.5	2684	9	AK056896	AK056896 Homo sapi
12	396	27.5	156534	2	AC009700	AC009700 Homo sapi
13	356.4	24.7	1536	6	AX405919	AX405919 Sequence
14	356.4	24.7	1923	9	BC029819	BC029819 Homo sapi
15	348.8	24.2	1521	10	BC019755	BC019755 Mus muscu
16	299.4	20.8	2439	9	AK027187	AK027187 Homo sapi
17	245.8	17.1	1706	9	BC020841	BC020841 Homo sapi
18	245.8	17.1	2684	6	AX714511	AX714511 Sequence
19	245.8	17.1	2684	9	AK056896	AK056896 Homo sapi
20	217	15.1	217	9	HS183H12F	Z57552 H. sapiens C
21	199.8	13.9	218	9	HS183H12F	Z57552 H. sapiens C
22	167.8	11.6	229583	10	AL844566	AL844566 Mouse DNA
23	166	11.5	1923	9	BC029819	BC029819 Homo sapi
24	165.8	11.5	231868	2	AC118124	AC118124 Rattus no
25	159.4	11.1	166937	2	AC012255	AC012255 Homo sapi
26	159.4	11.1	171444	2	AC087790	AC087790 Homo sapi
27	159.4	11.1	181312	9	AC091117	AC091117 Homo sapi
28	154.2	10.7	231868	2	AC118124	AC118124 Rattus no
29	147	10.2	229583	10	AL844566	AL844566 Mouse DNA
30	123.2	8.5	110250	2	AC138220	AC138220 Mus muscu
31	121.6	8.4	55061	2	AC091597	AC091597 Mus muscu
32	101	7.0	110250	2	AC138220	AC138220 Mus muscu
33	65.6	4.6	2207	3	AF429315	AF429315 Homo sapi
34	63.4	4.4	125020	9	AF429315	AF429315 Homo sapi
35	61	4.2	7218	6	I66494	I66494 Sequence 14
36	60.6	4.2	2000	6	AX655393	AX655393 Sequence
37	59.6	4.1	125020	9	AF429315	AF429315 Homo sapi
38	56.8	3.9	63689	2	AC024268	AC024268 Homo sapi
39	55.8	3.9	4615	1	STMWHB12X	L22864 Streptomyce
40	55.2	3.8	151173	8	AC135257	AC135257 Genomic s
41	53	3.7	2690	8	AY332473	AY332473 Oryza sat
42	53	3.7	169162	8	AP004267	AP004267 Oryza sat
43	52	3.6	2000	6	AX655393	AX655393 Sequence
44	52	3.6	117787	2	AC104271	AC104271 Oryza sat
45	51.8	3.6	168866	8	OSJN00023	AL606588 Oryza sat

ALIGNMENTS

RESULT 1
LOCUS AX035346 1441 bp DNA linear PAT 15-NOV-2000
DEFINITION Sequence 1 from Patent WO0053748.
ACCESSION AX035346
VERSION AX035346.1 GI:11191064
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Bruck,C.E., Vinals,Y.D., Coche,T. and Cassart,J.P.
TITLE Novel compounds
JOURNAL Patent: WO 0053748-A 1 14-SEP-2000;

BRUCK CLAUDINE ELVIRE MARIE (BE) ; SMITHKLINE BEECHAM BIOLOG (BE) ;
VINALS Y DE BASSOLS CARLOTA (BE) ; COCHE THIERRY (BE) ; CASSART
JEAN POL (BE)

FEATURES

```
1. .1441
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
```

ORIGIN

Query Match	100.0%;	Score 1441;	DB 6;	Length 1441;
Best Local Similarity	100.0%;	Pred. No. 3.3e-294;		
Matches 1441; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

OY		1	AAAGTAACGGCTACAGACAGTGAAGAAATAGTTTCGCTCGCCGGCTAAGAAAACTCTGTGC	60
Db		1	AAAGTAACGGCTACAGACAGTGAAGAAATAGTTTCGCTCGCCGGCTAAGAAAACTCTGTGC	60
OY		61	GTAACCAACCCAGAGCGCTTAGAGAGACGCCACTCCACGCTTCCTTAACGAGAGAGTGCA	120
Db		61	GTAACCAACCCAGAGCGCTTAGAGAGAGCCCACTCCACGCTTCCTTAACGAGAGAGTGCA	120
OY		121	GGACTCAGAATTCAACGAGCCCCACTCGGTCCAGCCTTGATACGCAAAGAGACGCCAAGGAC	180
Db		121	GGACTCAGAATTCAACGAGCCCCACTCGGTCCAGCCTTGATACGCAAAGAGACGCCAAGGAC	180
OY		181	GGCTCTCCCGCGTCCAGGACGCCCAAGCTTGCTGAGCTTGCCCCCGCTGCGTGACG	240
Db		181	GGCTCTCCCGCGTCCAGGACGCCCAAGCTTGCTGAGCTTGCCCCCGCTGCGTGACG	240
OY		241	ACTCGGCCGCGGTGCAGCATGACCCTGTGGAACGGCGTACTGCTTTTAACCCCAAGCCC	300
Db		241	ACTCGGCCGCGGTGCAGCATGACCCTGTGGAACGGCGTACTGCTTTTAACCCCAAGCCC	300
OY		301	CGGCATGCCGACAGGCTTCAAGCCTTCCAAGCTGCTCATCGTTATTTCTAGTGTTTTTGGCTCTA	360
Db		301	CGGCATGCCGACAGGCTTCAAGCCTTCCAAGCTGCTCATCGTTATTTCTAGTGTTTTTGGCTCTA	360
OY		361	GCAGCAAGCTTCTCTGCTCATCTTGCCGGGGATCCGTGGCCACTCCGCGTGGTTTGTTG	420
Db		361	GCAGCAAGCTTCTCTGCTCATCTTGCCGGGGATCCGTGGCCACTCCGCGTGGTTTGTTG	420
OY		421	GTGAGAGTTCTTCTCAGTCTGTTCTATAGGCGCAGAAATGTGGCTGTGACTTCAGTGCA	480
Db		421	GTGAGAGTTCTTCTCAGTCTGTTCTATAGGCGCAGAAATGTGGCTGTGACTTCAGTGCA	480
OY		481	GAATGTTTCGTTGGGTACAGTGAACAACAACATCCTAACAAAGCCTTCAGCGCACGCGC	540
Db		481	GAATGTTTCGTTGGGTACAGTGAACAACAACATCCTAACAAAGCCTTCAGCGCACGCGC	540
OY		541	GTTACAGCCCGTGTGCTGTCTCGTGGGCTGTGAGGGCATTATAATTACACTCACAGGG	600
Db		541	GTTACAGCCCGTGTGCTGTCTCGTGGGCTGTGAGGGCATTATAATTACACTCACAGGG	600
OY		601	ACCCAGTGCATCAGCTGAACGAGACCATTTGACTACAAAGAGCAGTTCACTTGGCGTCTG	660
Db		601	ACCCAGTGCATCAGCTGAACGAGACCATTTGACTACAAAGAGCAGTTCACTTGGCGTCTG	660
OY		661	AAAAGAAATTACGCCCGCGGAGTACGCCAAGCACTTGAGAAAGGGGCTGCCGAGCCAGTG	720
Db		661	AAAAGAAATTACGCCCGCGGAGTACGCCAAGCACTTGAGAAAGGGGCTGCCGAGCCAGTG	720
OY		721	CTCTAAGTGGCGGAGAGAGTTCAACACCGAGTAGCCCTTGCGGCTGTATACCAACAGTACAC	780
Db		721	CTCTAAGTGGCGGAGAGAGTTCAACACCGAGTAGCCCTTGCGGCTGTATACCAACAGTACAC	780
OY		781	CTGGCGGGACACTACGCTCGGGCAAGCTATGGGTGGCGTTCTGCTTCTGCTCCTCTCC	840
Db		781	CTGGCGGGACACTACGCTCGGGCAAGCTATGGGTGGCGTTCTGCTTCTGCTCCTCTCC	840
OY		841	AACGTGCTGCTCTCCAGCGCGGCCCCGCTCTACGAGGCTTGCACTGCTGACCAACCGGA	900
Db		841	AACGTGCTGCTCTCCAGCGCGGCCCCGCTCTACGAGGCTTGCACTGCTGACCAACCGGA	900

QY	901	GCCTTCGGGCTCTTCGGGGTCTTCGCTTGGCTTCATCTCTAGCGTGGCGCTCTGCCCC	960
Db	901	GCCTTCGGGCTCTTCGGGGTCTTCGCTTGGCTTCATCTCTAGCGTGGCGCTCTGCCCC	960
QY	961	CTCCGCTTAGGCTCCTCCGCGCTCACCACTCAGTACGGCGCCGCTTCTGGGTACCGCTG	1020
Db	961	CTCCGCTTAGGCTCCTCCGCGCTCACCACTCAGTACGGCGCCGCTTCTGGGTACCGCTG	1020
QY	1021	GCAACCGGCGTCTGTGCTCTTCTCCGAGGGGGCCGTGTGAGTCTCCAGTATGTTCCG	1080
Db	1021	GCAACCGGCGTCTGTGCTCTTCTCCGAGGGGGCCGTGTGAGTCTCCAGTATGTTCCG	1080
QY	1081	CCCAGCGCTTTCGCACCTTCTTGACCAAGCCCAAGACTGCAGCCAGAGAGAGG	1140
Db	1081	CCCAGCGCTTTCGCACCTTCTTGACCAAGCCCAAGACTGCAGCCAGAGAGAGG	1140
QY	1141	GGCTACCTCTTATCCTCGGCGAACCCACTGCACAAGCAGGCGCTCTCCAGACTTAAAA	1200
Db	1141	GGCTACCTCTTATCCTCGGCGAACCCACTGCACAAGCAGGCGCTCTCCAGACTTAAAA	1200
QY	1201	TGTATCACCACTAACCTGTGAGGGGAGACCAATCTGACTCTTCCCGCTTGGACAT	1260
Db	1201	TGTATCACCACTAACCTGTGAGGGGAGACCAATCTGACTCTTCCCGCTTGGACAT	1260
QY	1261	CGCAGGCGGGAGCAGTGCCTCGCAGGCTTGGGCCAGAGAGCTCCAGAAAGGCACTG	1320
Db	1261	CGCAGGCGGGAGCAGTGCCTCGCAGGCTTGGGCCAGAGAGCTCCAGAAAGGCACTG	1320
QY	1321	AGCGCTGCTGGCGCGAGGCTTCGACATCCGCAAGCAGGAAAGTCTCTGGGGCGA	1380
Db	1321	AGCGCTGCTGGCGCGAGGCTTCGACATCCGCAAGCAGGAAAGTCTCTGGGGCGA	1380
QY	1381	TCTGTAAATAAACCTTTTTCCTTTGTTTTTAAAAA	1440
Db	1381	TCTGTAAATAAACCTTTTTCCTTTGTTTTTAAAAA	1440
QY	1441	A 1441	
Db	1441	A 1441	

RESULT 2

FS7806241					
LOCUS	HSM806241	1740 bp	mRNA	linear	PRI 17-JUN-2003
DEFINITION	Homo sapiens mRNA; cDNA DKFZp686C04213 (from clone DKFZp686C04213);				
ACCESSION	EX537581				
VERSION	EX537581.1	GI:31873659			

SOURCE

ORGANISM	REFERENCE
Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	1 (bases 1 to 1740)

AUTHORS

Fobo, G., Han, M. and Wiemann, S.

TITLE

JOURNAL Submitted (17-JUN-2003) MIPPS, Ingolstaedter Landstr.1, D-85764

COMMENT

COMMENT	FEATURES
<p>neuroscience, Germany</p> <p>Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;</p> <p>sequenced by GBR (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.</p> <p>This clone (DKFZp686C04013) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cDNA/.</p>	<p>Location/Qualifiers</p>

FEATURES

```
source 1..1740
/organism="Homo sapiens"
/mol_type="mRNA"
```

```

/db_xref="taxon:9606"
/map="15q15"
/clone="DKFZp686C04213"
/tissue_type="human rectum tumor"
/clone_id="686 (synonym: hicc3). Vector pSPORT1_Sfi; host
DH10B; sites SfiI + SfiIB"
/dev stage="adult"
1694.1699
polya_signal
polya_site 1720

```

ORIGIN

```

Query Match      71.4%; Score 1029; DB 9; Length 1740;
Best Local Similarity 98.6%; Pred. No. 4.5e-207;
Matches 1038; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```

```

QY 381 CTTGCCGGGATCCGTGCGCACTCGCGCTGTTGGTTGGTGAAGTTCCTTCAGTCT 440
Db 688 CTGACCCGGGTGCTATTCTTCGACGCGCTGTTGGTTGGTGAAGTTCCTTCAGTCT 747
QY 441 GTTCATAGCGCAGAAATTGTGGCTGTGCACTTCAGTGCAGAAATGTTGGGTACAGT 500
Db 748 GTTCATAGCGCAGAAATTGTGGCTGTGCACTTCAGTGCAGAAATGTTGGGTACAGT 807
QY 501 GAACACCAACAACATCTTACAAAGCCTTCAGCGCAGCGCGGCTTACAGCCCGTTCGGTCT 560
Db 808 GAACACCAACAACATCTTACAAAGCCTTCAGCGCAGCGCGGCTTACAGCCCGTTCGGTCT 867
QY 561 GCTCGTGGGCTTGAGGGGCTTATATATACACTCAGAGGAGCCCGCATCAGCTGAA 620
Db 868 GCTCGTGGGCTTGAGGGGCTTATATATACACTCAGAGGAGCCCGCATCAGCTGAA 927
QY 621 CGAGACCATTAAGTACACAGCAGCAGTTCACCTGGCGTGAAGAGAAATTACGCCCGCGGA 680
Db 928 CGAGACCATTAAGTACACAGCAGCAGTTCACCTGGCGTGAAGAGAAATTACGCCCGCGGA 987
QY 681 GTACGGGAACGCACTGGAGAAAGGGGCTGCCGGAACCCAGTGTCTTACTGGCGGAGAAATT 740
Db 988 GTACGGGAACGCACTGGAGAAAGGGGCTGCCGGAACCCAGTGTCTTACTGGCGGAGAAATT 1047
QY 741 CACACCGAGTAGCCCTTGCGGCTGTACCAACAGTACCACTGGCGGAGACACTACGCCCTC 800
Db 1048 CACACCGAGTAGCCCTTGCGGCTGTACCAACAGTACCACTGGCGGAGACACTACGCCCTC 1107
QY 801 GGCACAGCTATGGGTGGCGTCTGCTTGGGCTCTCTCCACGCTGCTCCACGCC 860
Db 1108 GGCACAGCTATGGGTGGCGTCTGCTTGGGCTCTCTCCACGCTGCTCCACGCC 1167
QY 861 GGCACAGCTATGGGTGGCGTCTGCTTGGGCTCTCTCCACGCTGCTCCACGCC 920
Db 1168 GGCACAGCTATGGGTGGCGTCTGCTTGGGCTCTCTCCACGCTGCTCCACGCC 1227
QY 921 CTTGCCCTTGCCCTCCATCTTACAGCGTGGCGCTCTGCCGCTCCGCTAGGCTCCTCCGC 980
Db 1228 CTTGCCCTTGCCCTCCATCTTACAGCGTGGCGCTCTGCCGCTCCGCTAGGCTCCTCCGC 1287
QY 981 GCTACCACTCAGTACGCGCGGCTTCTGGGTACCGTGGCAACCGCGGCTCCTGTGCT 1040
Db 1288 GCTACCACTCAGTACGCGCGGCTTCTGGGTACCGTGGCAACCGCGGCTCCTGTGCT 1347
QY 1041 CTTCTCTGGAGGGCGGTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTCCACCTT 1100
Db 1348 CTTCTCTGGAGGGCGGTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTCCACCTT 1407
QY 1101 TCTGGAACCAAGCGGCAAGGACTGACAGCCAGAGAGAGAGGGGCTCACTCTTATCTCGG 1160
Db 1408 TCTGGAACCAAGCGGCAAGGACTGACAGCCAGAGAGAGAGGGGCTCACTCTTATCTCGG 1467
QY 1161 CGACCACTGACAAAGAGCGCGGCTCTCCAGACTTAAATGATATACCACTAACCTGTG 1220
Db 1468 CGACCACTGACAAAGAGCGCGGCTCTCCAGACTTAAATGATATACCACTAACCTGTG 1527
QY 1221 AGGGGACCAATCTGACTCCTTCCCGGCTTGGAGATCGCAGCGCGGAGACAGTGC 1280

```

```

Db 1528 AGGGGACCAATCTGACTCCTTCCCGGCTTGGAGATCGCAGCGCGGAGACAGTGC 1587
QY 1281 CCGCCAGGCTTGCGCCAGAGAGCTCCAGAGAAAGGCACTGAGCGCTGCGCGGAGGCC 1340
Db 1588 CCGCCAGGCTTGCGCCAGAGAGCTCCAGAGAAAGGCACTGAGCGCTGCGCGGAGGCC 1647
QY 1341 TCGGACATCCGAGGACCAAGGAAAGTCTCTGCGGCGATCTGTAAATAAACCTTTT 1400
Db 1648 TCGGACATCCGAGGACCAAGGAAAGTCTCTGCGGCGATCTGTAAATAAACCTTTT 1707
QY 1401 TCTTTTGTTTTAAAAA 1433
Db 1708 TCTTTTGTTTTAAAAA 1740

```

```

RESULT 3
BC031111
LOCUS
DEFINITION
Mus musculus RIKEN cDNA 9030623N16 gene, mRNA (cDNA clone MGC:35715
IMAGE:4989815), complete cds.
ACCESSION
BC031111
VERSION
BC031111.1 GI:21411407
KEYWORDS
MGC.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
REFERENCE
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schaeetz, T.E., Brownstein, M.J., Udell, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramsen, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalys, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

```

```

JOURNAL
MEDLINE
PUBMED
22388257
12477932
2 (bases 1 to 1354)
Strausberg, R.
Direct Submission
Submitted (03-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

```

REMARK
COMMENT

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov

FEATURES		5	rue JB Clement, Chatenay-Malabry 92296, France
source	location/Qualifiers	1..5851	
gene	/organism="Sus scrofa"		
	/mol_type="mRNA"		
	/db_xref="taxon:9823"	1..5851	
CDS	/gene="Duox1"	948..5609	
	/EC_number="1.6.3"		
	/function="putatively generates H2O2 by transferring electrons from NADH/NADPH to molecular oxygen"		
	/function="oxidoreductase acting on NAD(P)H with oxygen as acceptor"		
	/note="NADPH:O2 oxidoreductase"		
	/codon_start=1		
	/product="dual oxidase 1"		
	/protein_id="AA39338.1"		
	/db_xref="GI:23664373"		
	/translation="MGFRLALMTLLVGPMVPMGARNISWEVQRFDMYNNLMHKM GSKGRLQRLVPAYADGVYPLGEPHLPNPRLDLSNTAMGAPAGASLRNRTVLGVE GYHVLSDLVSIKPGCPAEFLNHIIPGDPVFDPHKSGDVLFPQSRMDPNTGQSPS NPRDLTNEVTGWLDSAIYGSBSWSDELRSFGQLASGDDPAFPQADPLFMWTP PDPATQGRPQGLAFYGAEGNEPFLQALGLMFRYHNLCAQKLAREHPLMGDELF QHARKVIATYOSITMYEWLPSLRKMPQETGYRPLDLSISPEFLAASEOFFSTMY PGVYRNASCHPOGVINRNSYSRALVCNSYMSREHNPLOAEVDVALLGMAST AEREDHMYEDVDQPMWPGPLKPSRTDHLASCLQGRDGLPSYTKARALGLPVTBW QDINPALSRSDDGLVLEATAALYNQDLSRLLELPGGLBESYGDGPLESTIVLDQFVRL RDGDRYWEENTKNGLESEKEIAEIRNTSLRDVLVAVNTMTPGALQPNVFMHAGDPCP QPRQSTKDLPAACPLIMRDYRKGSGFGVITIGLCCFPLVSLSAMVAQLRRNF KRLQVONROSIMCEKLVGKMLNEWQGRKEPCRPVLVHDSGOIHVMGRSLSVRTIQ LRPQOVNLILSNHGRRTLLKIPEYDVLMPDLSEERQVMENTLQSALESGLSE QEWLEQELMRAVTRQORSHLETFFRLFSQVLDIDODAGALPLDSSQKREAL TCELSRAEFAESLGLKQDMFYASMSPLADKNGYLSFEFLDILVFMKGSPEEKS RLMFMYDFDNGNLSKDEIFRMLRSFIEISNNCLSKALTEVESMEPEAGFQDKOE LTWEDFHEMLRDHSELRFTQLCVKVEVEVYIKDLCRASYISOEKLCPSPRVSAHC PRSNVDEVELTPMKLQCPDTPDPOEIRRRFGKVTSPDLFTBAHEKEFORSRH QTVQCFKRFVENYRRHIGLAVYTIAGGLFLERAYYAPAAHMGITDTRVGILIS RGTAASISFMFSYLLTMCRNLITPLRETFLRVYVPDAADVDFHRLIAGTAIILTVH SAGYVNVYLFSLBSLVSLCPGLFHDNGSEFPQKYVMWFQTVPGULVMLLIL AIMYVFASHHFRCSFRGFWLTHLILYLLYMLLIHGSFALIQLPRFIPLVLPALIV VGDKVLSLRKVEISVKAELLBSGVTHLQFORPQGFYKSGOWRIACLAGTTEY HPFTLSAPHEDTSLHIRAAAGWTRTLREISYSPPTDNCAPKLYLDGPFEGEHOE WHKEVSVLVGGGIVTPFASISLKLDFKSSVSCQVFCIKYFIWTRQORQFEMAD IIRREVENDRDLVSVHIYITQAEKFDLRTMLYICERHFQKVLNRSFTGLRSITH FGRPPEPFPNSLQEVHPQVRKIGVFCGSPGPMTKNVEKAQOLLINQDRTHFSHYEN F"		
ORIGIN			
Query Match	43.5%; Score 626.2; DB 4; Length 5851;		
Best Local Similarity	81.8%; Pred. No. 7.6e-122;		
Matches	735; Conservative 0; Mismatches 163; Indels 1; Gaps 1;		
QY	254	GCAGCATGACCTGTGGAACGGGCTACTGCTTTTACCAGCCCGGCGATGCGCAG	313
DB	10	GCAGCATGACTGTGTGAACGGGCTGCTGCTTCTACCTCAGCCTGGTATGCCGCCA	69
QY	314	GCTTCAGCGTTCACTGCTCATCGTTATTCTAGTGTGCTTGGCTCTAGACCAAGCTTCC	373
DB	70	GCTTCAGCGTCCCACTACTCATCGTTATTCTGTGTTCTTGGCCTTAGCCACAGCTTTC	129
QY	374	TGCTCATCTTGGCCGGGATCCGTGGCCACTGCGCTGTTTGGTTGTGAGAGTTCTTC	433
DB	130	TACTCATCTTGGCCGGGATTCGTGGCCACTGCGCTGTTCTGTGTTGTGAGAGTTCTTC	189
QY	434	TCACTCTGTTCATAGGCGCAGAAATTGTGCTGTGACCTTCAGTGCAGATGTTCTGG	493
DB	190	TCACTCTGTTCATAGGCGCAGAAATTGTGCGGTGACCTTCAGCGCAGATGACCGTGG	249
QY	494	GTAAGTGAACACCAACATCTTCAAAAGCCTTCAGCGCAGCGCGCTTACAGCCCGTG	553
DB	250	GCGAAGTTAGCAGCAATACATCTTCAAGGCTTCAAGTGTGCGACGTTCCAAAGCCCA	309

QY	554	TCGGTCTGCTCGTGGGCGCTGAGGGCATTATATTACACTCACAGGAGCCCACTGCATC	613
DB	310	TCGGTCTGCACGTGGGCGCTGAGGGGTGTGATATATACACTCACAGGGAATCCGCTAGGC	369
QY	614	AGCTGAACGAGACCATTTGACTACACAGAGCACTTACCTGGCGCTGAAAGAAATTACG	673
DB	370	AGCTGAATGAGACCATTCGACTACACAGAGCACTTACCTGGCGGTATGGGGAATACTATG	429
QY	674	CCGCGAGTACGCGAACGCACTGAGAGAGGGGCTGCCGAGCCAGTGTCTACTAGTGGCG	733
DB	430	CGCGCGCTATGCGAAGGCTCTGAGAGAGGGGCTGCCGAGCCAGTGTCTACTAGTGGCG	489
QY	734	AGAAGTTACACCGAGTAGCCCTTGGCGCTGTACACACAGTACCACTGGCGGACACT	793
DB	490	AGAAGTT-ACGCCAGAAAGCCCTGCGGGTTTACCGCAATATCGCCTTAGCCGACACT	548
QY	794	ACGCTCGGCGCAGCTATGAGTGGCGCTTCTGCTTCTGCTCTCTCCAACTGCTGCTCT	853
DB	549	ACGCTCGGCGCAGCTATGAGTGGCGCTTCTGCTTCTGCTCTCTCCAACTGCTGCTCT	608
QY	854	CCAGCGCGCGCGCTCTAGCAGAGCGCTGGCACTGCTGACCAACCGAGCCTTGGCGCTCT	913
DB	609	CCATGCGCGTCCCGCATATGAGAGCGCTGCTGCTCTCTCATACACCGCGCTTGGCGCTCT	668
QY	914	TCGGGCTTCTGCGCTTGGCGCTGCTCTCTAGCGTGGCGCTGCTGCGCGCTGAGGCT	973
DB	669	TCTGCTTCTGCGCTTGGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	728
QY	974	CCTCGCGCTCACCACTCAGTACGAGCGCGCGCTTCTGCGTCAACCGTGGCAACCGCGCTCC	1033
DB	729	CCTCGCAACTCACCACTCAGTACGAGCGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCT	788
QY	1034	TGTGCTCTTCTCTGAGGGGCGCGTGTGAGTCTCCAGTATGTTGGCGCGCGCTCTTC	1093
DB	789	TGTGCTCTTCTCTGAGGAGCGGCTGTGAGTCTCTCACTACGCTGACCGCGCTCTTC	848
QY	1094	GCACCTTCTTGACCAAGCGCGCAAGACTGACGACGAGAGAGGGGCTTCACTCTT	1152
DB	849	GCACCTTCTTGAGGGAAGTATCAAGACTGCGATAGTCAGGCAAAAGGCTTCCATT	907
RESULT 5			
AX035348	498 bp	DNA	linear PAT 15-NOV-2000
LOCUS	AX035348		
DEFINITION	Sequence 3 from Patent WO0053748.		
ACCESSION	AX035348		
VERSION	AX035348.1 GI:11191065		
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Bruck,C.E., Vinals,Y.D., Coche,T. and Cassart,J.P.		
TITLE	Novel compounds		
JOURNAL	Patent: WO 0053748-A 3 14-SEP-2000; BRUCK CLAUDINE ELVIRE MARIE (BE) ; SMITHKLINE BEECHAM BIOLOG (BE) ; VINALS Y DE BASSOLS CARLOTTA (BE) ; COCHE THIERRY (BE) ; CASSART JEAN POL (BE)		
FEATURES			
source	location/Qualifiers	1..498	
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match	32.3%; Score 465.4; DB 6; Length 498;		
Best Local Similarity	99.4%; Pred. No. 6.6e-88;		
Matches	488; Conservative 0; Mismatches 1; Indels 2; Gaps 2;		
QY	939	CTCTAGCGTGGCGCTTGGCGGCT-CCGCTAGGCTCTCGCGCTCACCACCTAGTACG	997

Db 1 CTTAGCGTGCCGCTCTGCCCCGCTCCGCTAGGCTCCTCCGCGCTCACCACCTCAGTACG 60

QY 998 -GGCCGCTTCTGGTCAAGCTGGCAACCGCGCTCTGCTCTTCTCGAGGGCC 1056

Db 61 AGCGCGCTTCTGGTCAAGCTGGCAACCGCGCTCTGCTCTTCTCGAGGGCC 120

QY 1057 GTGATGATCTCCAGTATGTTCCGCCAGCGCTCTTCCACCTTCTGACCAAGCGCC 1116

Db 121 GTGTGATGCTCCAGTATGTTCCGCCAGCGCTCTTCCACCTTCTGACCAAGCGCC 180

QY 1117 AAGACTGCAGCCAGAGAGAGAGGGGCTCACCCTTATCTTCTCGGCGACCCACTGCACAAG 1176

Db 181 AAGACTGCAGCCAGAGAGAGAGGGGCTCACCCTTATCTTCTCGGCGACCCACTGCACAAG 240

QY 1177 CAGCGCGCTCTCCAGACTTAATATGATCACCCTTATCTTCTCGGCGACCCACTGCACAAG 1236

Db 241 CAGCGCGCTCTCCAGACTTAATATGATCACCCTTATCTTCTCGGCGACCCACTGCACAAG 300

QY 1237 GACTCCTTCTCCCGCTTGGGACATCGAGCGCGGGAAGCAGTCCCGCCAGGCTGGGCC 1296

Db 301 GACTCCTTCTCCCGCTTGGGACATCGAGCGCGGGAAGCAGTCCCGCCAGGCTGGGCC 360

QY 1297 AGGAGAGCTTCCAGAGAGGAGCAGTGGCGCTGCTGGCGGAGGCTCGGACATCCGAGGC 1356

Db 361 AGGAGAGCTTCCAGAGAGGAGCAGTGGCGCTGCTGGCGGAGGCTCGGACATCCGAGGC 420

QY 1357 ACCAGGGAAGCTCTCTGGGGCGATCTGTAATAAACCTTTTCTTTTCTTTTAA 1416

Db 421 ACCAGGGAAGCTCTCTGGGGCGATCTGTAATAAACCTTTTCTTTTCTTTTAA 480

QY 1417 AAAAAAAAAA 1427

Db 481 AAAAAAAAAA 491

RESULT 6

LOCUS AC012255 166937 bp DNA linear HTG 06-MAY-2001

DEFINITION Homo sapiens chromosome 15 clone RP11-109D20 map 15, WORKING DRAFT

AC012255

SEQUENCE, 22 unordered pieces.

AC012255

AC012255.4 GI:7684440

HTG; HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

TITLE 1 (bases 1 to 166937)

JOURNAL Homo sapiens chromosome 15, clone RP11-109D20

REFERENCE

AUTHORS 2 (bases 1 to 166937)

TITLE Unpublished

JOURNAL

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tittrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE Direct Submission

JOURNAL Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On May 3, 2000 this sequence version replaced gi:7209928. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L3737

Center clone name: 109 D 20

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 153806 bases at least Q40

Consensus quality: 160606 bases at least Q30

Consensus quality: 163275 bases at least Q20

Insert size: 164837; sum-of-contigs

Quality coverage: 4.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 22 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 2087: contig of 2087 bp in length

* 2088 2187: gap of 100 bp

* 2188 3626: contig of 1439 bp in length

* 3627 3726: gap of 100 bp

* 3727 4845: contig of 1119 bp in length

* 4846 4945: gap of 100 bp

* 4946 6702: contig of 1757 bp in length

* 6703 6802: gap of 100 bp

* 6803 8326: contig of 1524 bp in length

* 8327 8426: gap of 100 bp

* 8427 10494: contig of 2068 bp in length

* 10495 10594: gap of 100 bp

* 10595 13215: contig of 2621 bp in length

* 13216 13315: gap of 100 bp

* 13316 16451: contig of 3136 bp in length

* 16452 16551: gap of 100 bp

* 16552 20843: contig of 4292 bp in length

* 20844 20943: gap of 100 bp

* 20944 24940: contig of 3997 bp in length

* 24941 25040: gap of 100 bp

* 25041 29901: contig of 4861 bp in length

* 29902 30001: gap of 100 bp

* 30002 34189: contig of 4188 bp in length

* 34190 34289: gap of 100 bp

* 34290 38386: contig of 4097 bp in length

* 38387 38486: gap of 100 bp

* 38487 45827: contig of 7341 bp in length

* 45828 45927: gap of 100 bp

* 45928 51729: contig of 5802 bp in length

* 51730 51829: gap of 100 bp

* 51830 60123: contig of 8294 bp in length

* 60124 60223: gap of 100 bp

* 60224 73532: contig of 13309 bp in length

* 73533 73632: gap of 100 bp

* 73633 87433: contig of 13801 bp in length

* 87434 87533: gap of 100 bp

* 87534 101708: contig of 14175 bp in length

* 101709 101808: gap of 100 bp

* 101809 121193: contig of 19385 bp in length

* 121194 121293: gap of 100 bp

* 121294 139051: contig of 17758 bp in length

* 139052 139151: gap of 100 bp

* 139152 166937: contig of 27786 bp in length.

Location/Qualifiers

1..166937

/organism="Homo sapiens"

/mol_type="genomic DNA"

FEATURES

source

```

/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-109D20"
/clone_lib="RPC1-11 Human Male BAC"
1. .2087
/misc_feature
/note="assembly_fragment"
2188. .3626
/note="assembly_fragment"
3727. .4845
/note="assembly_fragment"
4946. .6702
/note="assembly_fragment"
6803. .8326
/note="assembly_fragment"
8427. .10494
/note="assembly_fragment"
10595. .13215
/note="assembly_fragment"
13316. .16451
/note="assembly_fragment"
16552. .20843
/note="assembly_fragment"
20944. .24940
/note="assembly_fragment"
25041. .29901
/note="assembly_fragment"
30002. .34189
/note="assembly_fragment"
34290. .38386
/note="assembly_fragment"
38487. .45827
/note="assembly_fragment"
45928. .51729
/note="assembly_fragment"
51830. .60123
/note="assembly_fragment"
60224. .73532
/note="assembly_fragment"
73633. .87433
/note="assembly_fragment
clone_end:SP6
vector_side:right"
87534. .101708
/note="assembly_fragment"
101809. .121193
/note="assembly_fragment"
121294. .139051
/note="assembly_fragment"
139152. .166937
/note="assembly_fragment
clone_end:T7
vector_side:right"

```

ORIGIN

Query Match	28.1%;	Score 405.4;	DB 2;	Length 166937;
Best Local Similarity	98.6%;	Pred. No. 4.7e-75;		
Matches 409;	Conservative	0;	Mismatches 6;	Indels 0;
				Gaps 0;

QY	1	AAAGTAAACGGCTACAGACAGTGAATAATAGTTTCGCTCGCCGGCTAGAAAAA	CTGTGC	60
Db	157716	AAAGTAAACGGCTACAGACAGTGAATAATAGTTTCGCTCGCCGGCTAGAAAAA	CTGTGC	157775
QY	61	GTACCAACCCCGAGCGCTTGAGAGACGCCACCTCCACGCTTCCTTAACGAGAGGTGCA		120
Db	157776	GTACCAACCCCGAGCGCTTGAGAGACGCCACCTCCACGCTTCCTTAACGAGAGGTGCA		157835
QY	121	GGACTCAGACTTCACCAAGCCCACTCGGTCCCAAGCCTTGTAACGCAAAAGACGCCAAAGAC		180
Db	157836	GGACTCAGACTTCACCAAGCCCACTCGGTCCCAAGCCTTGTAACGCAAAAGACGCCAAAGAC		157895
QY	181	GCGCTCTCCCGCGCTCCAGGCAAGCCCAAGCTTGCTGCTTGCCCTGCGGCTGCGTGAAC		240

Db	157896	GCGCTCTCCCGCGTCCAGGAGACCCCAAGCTTGCTGGCTTGCCCGGCTCGGTGCAGC	157953
QY	241	ACTGGCCGGCGTGACGATGACCCCTGTGGAACGGCGTACTGCTTTTACCCCAAGCC	300
Db	157956	ACTGGCCGGCGTGACGATGACCCCTGTGGAACGGCGTACTGCTTTTACCCCAAGCC	158015
QY	301	CGGCATGCCGAGGCTTCAGGGTTCCACTGCTCATCGTATTTCTAGTGTTTTGGCTCTA	360
Db	158016	CGGCATGCCGAGGCTTCAGGGTTCCACTGCTCATCGTATTTCTAGTGTTTTGGCTCTA	158075
QY	361	GCAGCAAGCTTCTGCTCATTTGGCCGGGATCCGTGGCCACTCGCGCTGTTTT	415
Db	158076	GCAGCAAGCTTCTGCTCATTTGGCCGGGATCCGTGGCCACTCGGTAAAGGTGT	158130

RESULT 7	AC091117	AC091117	181312 bp	DNA	linear	PRI 20-FEB-2002
LOCUS	AC091117					
DEFINITION		Homo sapiens chromosome 15 clone RP11-109D20 map 15q15, complete sequence.				
ACCESSION	AC091117	AC091117				
VERSION	AC091117.5	GI:18767395				
KEYWORDS	HTG.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 181312)
Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,

TITLE	Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 181312)
AUTHORS	Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S., Burke, U., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R., Pate, D. and Hood, L.
TITLE	Direct Submission
JOURNAL	Submitted (30-MAR-2001) Multimegabase Sequencing Center, Institute

REFERENCE	AUTHORS	TITLE
3 (bases 1 to 191312)	Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S., Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R., Pate, D. and Hood, L.	Direct Submission
		Submitted (13-FEB-2002) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA

REFERENCE	AUTHORS	TITLE
4 (bases 1 to 181312)	Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S., Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R., Pate, D. and Hood, L.	Direct Submission
		Submitted (14-FEB-2002) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA

REFERENCE 5 (bases 1 to 181312)
 AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
 Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
 Pate, D. and Hood, L.
 TITLE Direct Submission
 JOURNAL Submitted (20-FEB-2002) Multimegabase Sequencing Center, Institute
 for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
 98105, USA
 COMMENT On Feb 20, 2002 this sequence version replaced gi:18653547.

----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UMWSC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leerowen@systembiology.org
----- Summary Statistics
Sequencing vector: PUC18; L08752

Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399

Note: Data from overlapping clone AC090888 [Drafting center: WIBR] was added for finishing. RP11-109D20 spans bases 1-174236 of this sequence. Bases 174237-179373 derive from clone RP11-276K9, AC087790 [Drafting center: WIBR] to establish the overlap with the next BAC in our tiling path. Bases 179373-181312 derive from finished clone CTD-2651B20, AC051619 [Drafting center: UWMSC] in order to give overlap required for determining long range contiguity.

FEATURES

source

Location/Qualifiers
1.181312

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q15"

/clone="RP11-109D20"

/clone_lib="RPCI human BAC library 11"

/note="This clone overlaps CTD-2014N11 AC090888 and RP11-276K9 AC087790. Data from overlapping BACs were added and the consensus sequence determined from RP11-109D20 to the extent possible."

misc_feature

1.89137
/note="overlap with CTD-2014N11, AC090888"

unsure

36107.36120
/note="low quality data"

misc_feature

40470.40820
/note="sequence data generated from subcloned PCR product"

misc_feature

53030.53480
/note="sequence data generated from subcloned PCR product"

unsure

86982.86983
/note="low quality data"

misc_feature

95610.96030
/note="sequence data generated from subcloned PCR product"

misc_feature

96830.97420
/note="sequence data generated from subcloned PCR product"

misc_feature

98260.98420
/note="Bingle clone coverage"

unsure

98850.98852
/note="low quality data"

unsure

100981.101003
/note="low quality data"

misc_feature

174237.179372
/note="overlap with RP11-276K9, AC087790. This sequence data was added to establish the overlap between the BACs in the tiling path"

misc_feature

179373.181312
/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

misc_feature

179373.181312
/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

misc_feature

179373.181312
/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

misc_feature

179373.181312
/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

misc_feature

179373.181312
/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

misc_feature

179373.181312
/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

misc_feature

179373.181312
/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

misc_feature

179373.181312
/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

misc_feature

179373.181312
/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

misc_feature

179373.181312
/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

misc_feature

179373.181312
/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

misc_feature

179373.181312
/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

misc_feature

179373.181312
/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

misc_feature

179373.181312
/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

misc_feature

179373.181312
/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

misc_feature

179373.181312
/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

misc_feature

179373.181312
/note="overlap with CTD-2651B20, AC051619. This data is added to give overlap for long range contiguity."

Db 165196 GCGCTCTCCCGCGCTCCAGGACGCCAGCTTGCTGCTTGCCCTGCCCCGCTGCTGACG 165255

Qy 241 ACTGCGCGCGCGTGCAGCATGACCTGTGGAACGGCGTACTGCTTTTACCCCGAGCCC 300

Db 165256 ACTCGCGCGCGGTGCAGCATGACCTGTGGAACGGCGTACTGCTTTTACCCCGAGCCC 165315

Qy 301 CGGATGCGCGCGAGGCTTACGCGTTCACCTGCTCATGCTATTCTAGTGTGCTCTA 360

Db 165316 CGGATGCGCGAGGCTTACGCGTTCACCTGCTCATGCTATTCTAGTGTGCTCTA 165375

Qy 361 GCAGCAAGCTTCTGCTCATCTTGGCGGGATCCGTGGCCACTCGCGCTGTTT 415

Db 165376 GCAGCAAGCTTCTGCTCATCTTGGCGGGATCCGTGGCCACTCGCGTAAAGGTGT 165430

RESULT 8

AC009700/c 156534 bp DNA linear HTG 20-APR-2000

LOCUS Homo sapiens chromosome 15 clone RP11-163P10 map 15, WORKING DRAFT

DEFINITION SEQUENCE, 15 unordered pieces.

AC009700 AC009700.4 GI:7622346

VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 156534)

Barren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 15, clone RP11-163P10

Unpublished

2 (bases 1 to 156534)

Barren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,N.,

Baker,J., Baldwin,J., Barna,N., Beckertly,R., Benn,J., Brown,A.,

Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,

Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,

Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,

Funk,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,

Haggs,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,

Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,

Marquis,N., McEwan,P., McGuirk,A., McKernan,K., McLaughlin,J.,

Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,

Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,

Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,

Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,

Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,

Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

Direct Submission

Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 20, 2000 this sequence version replaced gi:6056272.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: 163_P_10

Center clone name: 12336

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 108837 bases at least Q40

Consensus quality: 130230 bases at least Q30

Insert size: 157000; agarose-fp

Insert size: 155134; sum-of-contrigs

Quality coverage: 3.2 in Q20 bases; agarose-fp

Quality coverage: 3.3 in Q20 bases; sum-of-contrigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
* 1 1169: contig of 1169 bp in length
* 1170 1269: gap of 100 bp
* 1270 3335: contig of 2066 bp in length
* 3336 3435: gap of 100 bp
* 3436 5698: contig of 2263 bp in length
* 5699 5798: gap of 100 bp
* 5799 9216: contig of 3418 bp in length
* 9217 9316: gap of 100 bp
* 9317 15633: contig of 6317 bp in length
* 15634 15733: gap of 100 bp
* 15734 21746: contig of 6013 bp in length
* 21747 21846: gap of 100 bp
* 21847 30351: contig of 8505 bp in length
* 30352 30451: gap of 100 bp
* 30452 36509: contig of 6058 bp in length
* 36510 36610: gap of 100 bp
* 36610 45279: contig of 8670 bp in length
* 45280 45379: gap of 100 bp
* 45380 56851: contig of 11472 bp in length
* 56852 56951: gap of 100 bp
* 56952 67374: contig of 10423 bp in length
* 67375 67474: gap of 100 bp
* 67475 79534: contig of 12060 bp in length
* 79535 79634: gap of 100 bp
* 79635 95432: contig of 15798 bp in length
* 95433 95532: gap of 100 bp
* 95533 120801: contig of 25269 bp in length
* 120802 120901: gap of 100 bp
* 120902 156534: contig of 35633 bp in length.
```

FEATURES

Source

```
1. 156534
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-163P10"
1. 1169
/clone_lib="RPC1-11 Human Male BAC"
misc_feature
/feature="assembly_fragment"
1270..3335
misc_feature
/feature="assembly_fragment"
3436..5698
misc_feature
/feature="assembly_fragment"
5799..9216
misc_feature
/feature="assembly_fragment"
9317..15633
misc_feature
/feature="assembly_fragment"
15734..21746
misc_feature
/feature="assembly_fragment"
21847..30351
misc_feature
/feature="assembly_fragment"
30452..36509
/feature="assembly_fragment"
clone_end:T7
vector_side:left"
36610..45279
/feature="assembly_fragment"
45380..56851
/feature="assembly_fragment"
56952..67374
/feature="assembly_fragment"
67475..79534
/feature="assembly_fragment"
```

```
misc_feature 79635..95432
/feature="assembly_fragment"
misc_feature 95533..120801
/feature="assembly_fragment"
misc_feature 120902..156534
/feature="assembly_fragment"
```

ORIGIN

Query Match 28.0%; Score 403.8; DB 2; Length 156534;
Best Local Similarity 98.3%; Pred. No. 1e-74;
Matches 408; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
QY 1 AAAGTAAACGGCTACAGACAGATGAGAAATAGTTTCGCTCGCCGCTAGAAAACTCTGTG 60
DB 7148 AAAGTAAACGGCTACAGACAGATGAGAAATAGTTTCGCTCGCCGCTAGAAAACTCTGTG 7089
QY 61 GTACCAACCCAGAGCGTTGAGAGCAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 120
DB 7088 GTACCAACCCAGAGCGTTGAGAGCAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 7029
QY 121 GGACTCAGACTTACACCAAGCCCACTCGTCCAGCTTGTACGCAAGAGACGCCAAGAC 180
DB 7028 GGACTCAGACTTACACCAAGCCCACTCGTCCAGCTTGTACGCAAGAGACGCCAAGAC 6969
QY 181 GCGCTCTCCCGCGCCAGAGCCCAAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
DB 6968 GCGCTCTCCCGCGCCAGAGCCCAAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6909
QY 241 ACTCGCCCGCGCGTGCAGCATGACCTGTGGAACGGCGTACTGCTTTTACCCCAAGCCC 300
DB 6908 ACTCGCCCGCGCGTGCAGCATGACCTGTGGAACGGCGTACTGCTTTTACCCCAAGCCC 6849
QY 301 CGGCATCGCCGAGCGCTTACAGCGCTTCCACTGCTCATGCTTATTTAGTGTGCTCTA 360
DB 6848 CGGCATCGCCGAGCGCTTACAGCGCTTCCACTGCTCATGCTTATTTAGTGTGCTCTA 6789
QY 361 GCAGCAAGCTTCTGCTCATCTTGCAGCGGATCCGCTGCGCACTCGCGCTGTTT 415
DB 6788 GCAGCAAGCTTCTGCTCATCTTGCAGCGGATCCGCTGCGCACTCGGTAAAGGTGT 6734
```

RESULT 9

AC087790

LOCUS

DEFINITION

AC087790

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

ATTNORS

AC087790 171444 bp DNA linear HTG 09-MAY-2001
Homo sapiens chromosome 15 clone RP11-276K9 map 15, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
AC087790 GI:13357344
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULITOP.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 171444)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-276K9
Unpublished
2 (bases 1 to 171444)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Labocque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,

TITLE
JOURNAL
COMMENT

Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

Submitted (24-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2001 this sequence version replaced gi:12408501.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L12451

Center clone name: 276_K_9

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 165906 bases at least Q40
Consensus quality: 168587 bases at least Q30
Consensus quality: 169557 bases at least Q20
Insert size: 18000; agarose-efp
Insert size: 170244; sum-of-contigs
Quality coverage: 5.7 in Q20 bases; agarose-efp
Quality coverage: 6.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 6091: contig of 6091 bp in length
* 6092 6191: gap of 100 bp
* 6192 7986: contig of 1795 bp in length
* 7987 8086: gap of 100 bp
* 8087 43939: contig of 35853 bp in length
* 43940 44039: gap of 100 bp
* 44040 45298: contig of 1259 bp in length
* 45299 45398: gap of 100 bp
* 45399 47595: contig of 2197 bp in length
* 47596 47695: gap of 100 bp
* 47696 52385: contig of 4690 bp in length
* 52386 52485: gap of 100 bp
* 52486 56310: contig of 3825 bp in length
* 56311 56410: gap of 100 bp
* 56411 71313: contig of 14903 bp in length
* 71314 71413: gap of 100 bp
* 71414 90993: contig of 19580 bp in length
* 90994 91093: gap of 100 bp
* 91094 111459: contig of 20366 bp in length
* 111460 111559: gap of 100 bp
* 111560 137726: contig of 26167 bp in length
* 137727 137826: gap of 100 bp
* 137827 170053: contig of 32227 bp in length
* 170054 170153: gap of 100 bp
* 170154 171444: contig of 1291 bp in length.
```

FEATURES

source

```
1. 171444
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
```

```
/clone="RP11-276K9"
/clone_1lb="RP11 Human Male BAC"
1. 6091
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
6192..7986
/note="assembly_fragment"
8087..43939
/note="assembly_fragment"
44040..45298
/note="assembly_fragment"
45399..47595
/note="assembly_fragment"
47696..52385
/note="assembly_fragment"
52486..56310
/note="assembly_fragment"
56411..71313
/note="assembly_fragment"
71414..90993
/note="assembly_fragment"
91094..111459
/note="assembly_fragment"
111560..137726
/note="assembly_fragment"
137827..170053
/note="assembly_fragment"
170154..171444
/note="assembly_fragment"
clone_end:T7
vector_side:right"
```

ORIGIN

```
Query Match 28.0%; Score 403.8; DB 2; Length 171444;
Best Local Similarity 98.3%; Pred. No. 1e-74;
Matches 408; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 AAAGTACGGCTACAGACGTGAGAAATAGTTTCGCTCGCCGGCTAGAAAAACTCTGTG 60
Db 140413 AAAGTACGGCTACAGACGTGAGAAATAGTTTCGCTCGCCGGCTAGAAAAACTCTGTG 140472

QY 61 GTACCAACCCAGAGCGTTGAGAGCAGCCCACTCCAGCTTCTTAACGAGAGGTGCA 120
Db 140473 GTACCAACCCAGAGCGTTGAGAGCAGCCCACTCCAGCTTCTTAACGAGAGGTGCA 140532

QY 121 GGACTCAGACTTCACCAAGCCCACTCGTCCAGCTTGCTGTAAGCAAGAGAGCAAGAC 180
Db 140533 GGACTCAGACTTCACCAAGCCCACTCGTCCAGCTTGCTGTAAGCAAGAGAGCAAGAC 140592

QY 181 GGGCTCTCCGCGTCCAGCAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db 140593 GGGCTCTCCGCGTCCAGCAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 140652

QY 241 ACTCGCCGCGGTGAGCAATGACCTGTGGAAGCGGCTACTGCTTTTACCCCAAGCCC 300
Db 140653 ACTCGCCGCGGTGAGCAATGACCTGTGGAAGCGGCTACTGCTTTTACCCCAAGCCC 140712

QY 301 CGGATGCCGAGGCTTCACGCTTCACCTGCTCATCTGTTATTCTAGTGTTTTGGCTCTA 360
Db 140713 CGGATGCCGAGGCTTCACGCTTCACCTGCTCATCTGTTATTCTAGTGTTTTGGCTCTA 140772

QY 361 GCAGCAAGCTTCTGCTCATCTTGGCCGGGATCCGTGGCCACTCGCGCTGTTT 415
Db 140773 GCAGCAAGCTTCTGCTCATCTTGGCCGGGATCCGTGGCCACTCGCGGTAAGGTGT 140827
```

```
RESULT 10
AX714511/c AX714511 2684 bp DNA linear PAT 15-APR-2003
LOCUS AX714511
DEFINITION Sequence 1195 from Patent EP1293569.
ACCESSION AX714511
VERSION AX714511.1 GI:29889464
```



```

KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   1
AUTHORS     Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,
            Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,
            Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and
            Masuno, Y.
TITLE       Full-length cDNAs
JOURNAL     Patent: EP 1293569-A 1195 19-MAR-2003;
            Helix Research Institute (JP) ; Research Association for
            Biotechnology (JP)
FEATURES    source
            location/Qualifiers
            1..2684
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

ORIGIN
Query Match      27.5%; Score 396; DB 6; Length 2684;
Best Local Similarity 97.6%; Pred. No. 3.5e-73;
Matches 402; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1007 TCTGGGTACGCTGGCAACCGCGCTGCTGTGCTCTTCTCTCGAGAGGGCCGTGTAGTC 1066
DB 2363 TTTCGATCCCCACCGCCACAGGCGCTGCTGTGCTCTTCTCTCGAGAGGGCCGTGTAGTC 2304

QY 1067 TCCAGTATGTTGGGCCCGAGCGCTCTTTCGACCCCTTCTGACCAAGCGCCAGACTGCA 1126
DB 2303 TCCAGTATGTTGGGCCCGAGCGCTCTTTCGACCCCTTCTGACCAAGCGCCAGACTGCA 2244

QY 1127 GCCAGGACAGAGGGGCTCACTCTTATTCCTCGGCGGACCCACTGCAACAAGCGCCGCTC 1186
DB 2243 GCCAGGACAGAGGGGCTCACTCTTATTCCTCGGCGGACCCACTGCAACAAGCGCCGCTC 2184

QY 1187 TCCCAAGCTTAAATGTATCACCACCTAAGCTGTGAAGGGGAGCCCAATCTGACTCCTTC 1246
DB 2183 TCCCAAGCTTAAATGTATCACCACCTAAGCTGTGAAGGGGAGCCCAATCTGACTCCTTC 2124

QY 1247 CCGCCTTGGGACATCGAGGCGCGGGAAGCAGTGCCTCCGCGGAGGCTGGGCGGAGAGCTC 1306
DB 2123 CCGCCTTGGGACATCGAGGCGCGGGAAGCAGTGCCTCCGCGGAGGCTGGGCGGAGAGCTC 2064

QY 1307 CAGGAAGGGCACTGAGCGCTGTGCGCGAGGCGCTCGACATCCGACAGGACAGGAAA 1366
DB 2063 CAGGAAGGGCACTGAGCGCTGTGCGCGAGGCGCTCGACATCCGACAGGACAGGAAA 2004

QY 1367 GTCTCTGGGGCGATCTGTAATAAACCCTTTTCTTTGTTTTTAAAAA 1418
DB 2003 GTCTCTGGGGCGATCTGTAATAAACCCTTTTCTTTGTTTTTAAAAA 1952

RESULT 11
AK056896/c 2684 bp mRNA linear PRI 01-AUG-2002
LOCUS      Homo sapiens cDNA FLJ32334 f.5, clone PROST2005426.
ACCESSION AK056896
VERSION   AK056896.1 GI:16552419
KEYWORDS  oligo capping; f.5 (full insert sequence).
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE  1
AUTHORS   Ishibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S.,
            Ono, Y., Horiuchi, T., Hiraoka, S., Murakawa, K., Takiguchi, S.,
            Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M.,
            Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,
            Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y.,
            Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y.,
            Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A.,

```

```

TITLE       Unpublished
JOURNAL     NEDO human cDNA sequencing project
AUTHORS     Isogai, T., Otsuki, T. and Sugiyama, T.
REFERENCE   2 (bases 1 to 2684)
TITLE       Direct Submission
JOURNAL     Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: RAB and
            HRI.

FEATURES    source
            Location/Qualifiers
            1..2684
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="PROST2005426"
            /tissue_type="prostate"
            /clone_1fb="PROST2"
            /note="cloning vector: pME18SFL3"
            56..952
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="BAB71304.1"
            /db_xref="GI:16552420"
            /translation="MATLGHFPFPYAGPRPTFPMDTLASIIMIFLTALATFIVILPG
            IRKTRFLFWLRVTSLSFLGALIGTPVOQNETINNEEFTWRLGENYAEYAKALE
            KGLPDPVLYLAEKFTPRSPGLYRQYRLAGHYTSAMLTAFLCMLANVMSPLVLY
            GGMULATGIFOLLALIFSMATSLTSPCLHGLASVLAHTHGPAFTLLTGLCVL
            LGLMAVAHRMQPHRLKAFPNQSVDEDPMLWSPBEGGILSPRYRSMADSPKSQDPL
            SEASSTKAYCKEAKHPKDPDCAI"

ORIGIN
Query Match      27.5%; Score 396; DB 9; Length 2684;
Best Local Similarity 97.6%; Pred. No. 3.5e-73;
Matches 402; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1007 TCTGGGTACGCTGGCAACCGCGCTCTGTGCTCTTCTCTCGAGAGGGCCGTGTAGTC 1066
DB 2363 TTTCGATCCCCACCGCCACAGGCGCTCTGTGCTCTTCTCTCGAGAGGGCCGTGTAGTC 2304

QY 1067 TCCAGTATGTTGGGCCCGAGCGCTCTTGCACCCCTTTCGACCAAGCGCCAGACTGCA 1126
DB 2303 TCCAGTATGTTGGGCCCGAGCGCTCTTGCACCCCTTTCGACCAAGCGCCAGACTGCA 2244

QY 1127 GCCAGGACAGAGGGGCTCACTCTTATCTCGGCGGACCACTGACCAAGCAGGCGGCTC 1186
DB 2243 GCCAGGACAGAGGGGCTCACTCTTATCTCGGCGGACCACTGACCAAGCAGGCGGCTC 2184

QY 1187 TCCCAAGCTTAAATGTATCACCACCTAAGCTGTGAAGGGGAGCCCAATCTGACTCCTTC 1246
DB 2183 TCCCAAGCTTAAATGTATCACCACCTAAGCTGTGAAGGGGAGCCCAATCTGACTCCTTC 2124

QY 1247 CCGCCTTGGGACATCGAGGCGCGGGAAGCAGTGCCTCCGCGGAGGCTGGGCGGAGAGCTC 1306
DB 2123 CCGCCTTGGGACATCGAGGCGCGGGAAGCAGTGCCTCCGCGGAGGCTGGGCGGAGAGCTC 2064

QY 1307 CAGGAAGGGCACTGAGCGCTGTGCGCGAGGCGCTCGACATCCGACAGGACAGGAAA 1366
DB 2063 CAGGAAGGGCACTGAGCGCTGTGCGCGAGGCGCTCGACATCCGACAGGACAGGAAA 2004

QY 1367 GTCTCTGGGGCGATCTGTAATAAACCCTTTTCTTTGTTTTTAAAAA 1418
DB 2003 GTCTCTGGGGCGATCTGTAATAAACCCTTTTCTTTGTTTTTAAAAA 1952

```


QY	1067	TCCAGTATGTTGGGCCAGCGGCTCTTCGCACACCCTTCTGGACCAAAAGCCCAAGGACTGCA	1126
Db	96171	TCCAGTATGTTGGGCCAGCGGCTCTTCGCACACCCTTCTGGACCAAAAGCCCAAGGACTGCA	96230
QY	1127	GCCAGGAGAGAGGGGGCTCACTCTTATCTCTGGCGAACCCCACTGCACAAGCAGGCCGCTC	1186
Db	96231	GCCAGGAGAGAGGGGGCTCACTCTTATCTCTGGCGAACCCCACTGCACAAGCAGGCCGCTC	96290
QY	1187	TCCAGACTTAAATGTATCAACCACTAACCTGTGAGGGGAGCCCAATCTGACTCCTTCC	1246
Db	96291	TCCAGACTTAAATGTATCAACCACTAACCTGTGAGGGGAGCCCAATCTGACTCCTTCC	96350
QY	1247	CCGCTTGGGACATCGCAGGCCGGAAAGCAGTGCCCGCCAGGCTTGGCCAGAGAGCTC	1306
Db	96351	CCGCTTGGGACATCGCAGGCCGGAAAGCAGTGCCCGCCAGGCTTGGCCAGAGAGCTC	96410
QY	1307	CAGGAAGGGCACTGAGCGCTGTGGCGGAGGCGCTGGACATCCGCGAGGACCAAGGAAA	1366
Db	96411	CAGGAAGGGCACTGAGCGCTGTGGCGGAGGCGCTGGACATCCGCGAGGACCAAGGAAA	96470
QY	1367	GTCTCCTGGGGCGATCTGTAAATAAACCTTTTCTTTTGTGTTTTTAAAA	1418
Db	96471	GTCTCCTGGGGCGATCTGTAAATAAACCTTTTCTTTTGTGTTTTTAAAA	96522

RESULT 13	AX405919	LOCUS	AX405919	DEFINITION	Sequence 334 from Patent WO0222660.	1636 bp	DNA	linear	PAT 14-JUN-2002
VERSION	AX405919	ACCESSION	AX405919	KEYWORDS	AX405919.1	GI:21439277			
SOURCE		ORGANISM	Homo sapiens (human)		Homo sapiens				
REFERENCE		AUTHORS	Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F., Xue, A.J., Yang, Y., Wehrman, T. and Drmanac, R.T.						
TITLE		JOURNAL	Novel nucleic acids and polypeptides						
			Patent: WO 0222660-A 334 21-MAR-2002;						
			HYSEQ, INC. (US)						
FEATURES		source	Location/Qualifiers						
			1..1636						
			/organism="Homo sapiens"						
			/mol_type="unassigned DNA"						
			/db_xref="taxon:9606"						
			441..1472						

```

/note="unnamed protein product"
/codon_start=1
/protein_id="CAD35026.1"
/db_xref="GI:21439278"
/db_xref="REMTREMBL:CAD35026"
/translation="MATLGHTFPFVGPKPFPMDTTLASIMIPLYATATFIVILPG
IRGKTRFLMLRVVTSLSFIGAAILAVNFSSMSVGVSTNTSYKAFSSEWISADIGLQ
VGLGVNITLTKGTPVQQLNETINNEEFRTWLRGENVAEYAKALEGLPDPVLVTEK
FTPRSPCGLYROYRLAGHYTSAMLWVAFLCWLAVNMLSPALVYGYMLLATGTFQD
LALLFMSMATSLTSPCPPLHIGASVYHHHGPAFWITLTGGLCVLLGLAMAVAHNRQ
PRLKAFPNQSVDEDPMLMEWSPEEGGLSPRYRSMADSPKSDIPSEASSTKAYCKEA
HPKDPDCAL"

```

Query Match	24.7%	Score 356.4	DB 6	Length 1636
Best Local Similarity	63.9%	Pred. No. 8e-65		
Matches 556	Conservative	0	Mismatches 311	Indels 3
				Gaps 1
QY	255	CAGCATGACCCCTGTGGAACGGCGTACTGCCTTTTAAACCCCAAGCCCGGCATGCCGACAGG	314	
DB	437	CAGATGGCTACTTTGGGACACACATTCCTTATGCTGGCCCCAAGCCAACTTCCC	496	
QY	315	CTTCAGCGTTCCACTGCTCATGTTATTCTAGTGTTTTGGCTCTAGACGACGAAGCTTCCT	374	

Db	497	GATGACACCACTTTGGCCAGCATCATCATGATCTTTCTGACTGCACTGGCCACGTTCA	556
QY	375	GCTCATCTTGGCCGGGATCCGTGGCCACTCGCGCTGTTTGGTGGTGAAGTCTTCT	434
Db	557	CGTCATCTCGCTGGCATTCGGGGAAAGACGAGGCTGTCTCTGGCTCTCGGGTGTGAC	616
QY	435	CAGTCTGTTCA TAGCGCAGAAATTGTGCTGTGCATTCAGTGCAGATGGTTCTGGG	494
Db	617	CAGCTATTATCGGGGCTGCCATCTCGCTGTGAATTCAGTTCTGAGTGTCTGTGGG	676
QY	495	TACAGTGACCAACACACATCTCTACAAGCCTTCAGCGCAGCGCGCTTACAGCCCGT	554
Db	677	CCAGTCAAGACCAACACATCATACAGGCGCTTCAGTTCTGAGTGAATGATCAGCGCTGATAT	736
QY	555	CGGTCGTCTGTGGGCTTGAGGGCATTAATATTACATCA CAGGAGCCCACTGCATCA	614
Db	737	TGGGCTGCAAGTGGGCTGGGTGGAGTCAATCACTACACAGGAGCCCGCTGCAGCA	796
QY	615	GCTGAACGAGACCATTAACAACGAGCACTTCACTCGCGCTGTGAAGAATTAAGC	674
Db	797	GCTGAATGAGACCATTAACAACGAGAGTTCACTGGCGCTGGGTGAGAACTATGC	856
QY	675	CGCGAGTACCGGAACGCACTGGAGAGGGGCTCGCGGACCCAGTGTCTTACTGGCGGA	734
Db	857	TGAGGAGTATGCAAGGCTCTGGAGAGGGGCTGCCAGACCTGTGTGTACTTA	916
QY	735	GAAGTTCACACCGAGTAGCCCTTGGCGCTGTATCCACCAGTACCACCTGGCGGACACTA	794
Db	917	GAAGTTCATCCAAAGAGCCATGTGGCTTAATCCGCCAGTACCAGCTTGGCGGACACTA	976
QY	795	CGCCTCGGCCACGCTATGGGTGGCGTTCTGTCTTCTGGCTCTTCCAACTGTGCTCTC	854
Db	977	CACCTCAGCCATGCTATGGGTGGCATTCCTCTGCTGTGGCTGGCCATGTATGCTCTC	1036
QY	855	CACGCCGGCCCCGCTCTACGAGAGGCTGGCACTGCTGACCAACCGAGCCTTGGCGCTCT	914
Db	1037	CATGCTGCGTGTATATGTGTGGCTACATGCTATTGGCAACGGGCATCTTCCAGCTGTT	1096
QY	915	CG---GGGTCTTGGCCTTGGCCTCCATCTCTAGCGTGCAGCTTGGCCGCTCGCCTAGG	971
Db	1097	GGGTCTGCTCTTCTTCTCAATGGCCACATACACTCACCTGACCTGTCCCTGCACCTGGG	1156
QY	972	CTCCTCCGCGCTCACCACTCACTAGTACGCGCGCCTTCTGGGTCAAGCTGGCAACGGCGT	1031
Db	1157	CGCTTCTGTGCTCATACTCAACATGAGCGCTTCTGTGAATCAATGACCA CAGACT	1216
QY	1032	CTGTGCTCTTCTCTCGAGAGGCGCGTGTGAGTCTCCAGTATGTTCCGCGCCAGCGCTCT	1091
Db	1217	GCTGTGTGTGCTGTGGCGCTGTGCTATGGCGGTGGCCCAAGAGATGAGCTTCA CAGGCT	1276
QY	1092	TGCAACCCCTTCTGGAACCAAGCGCCAAAGGA	1121
Db	1277	GAAGGCTTCTTCAACCAAGATGTGATGA	1306

RESULT 14				
BC029819				
LOCUS				
DEFINITION		1923 bp	mRNA	linear PRI 06-OCT-2003
ACCESSION	BC029819			
VERSION	MGCL35310			
KEYWORDS	IMAGE:5177619,			complete cds.
SOURCE	BC029819.1	GI:20987583		
ORGANISM	MGC.			
	Homo sapiens (human)			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 1923)			
AUTHORS	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buettow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,			

VERSION BC019755.1 GI:18043594
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1521)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Boufard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.B., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 2238257
PUBMED 12477932
REFERENCE 2 (bases 1 to 1521)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louie, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Naravati, A.N., Gibbs, R.A.
FEATURES
SOURCE
1.1521
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:25550 IMAGE:3964798"
/tissue_type="Mammary tumor. Metastatic tumor. TGF alpha model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI CGAP_Mam1"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1.1521
/gene="BC019755"
/note="synonym: MGC25550"
/db_xref="LocusID:213696"

CDS
/db_xref="MGI:2384861"
285..1310
/product="CDNA sequence" BC019755
/protein_id="AAH19755.1"
/db_xref="GI:18043595"
/translation="MAALGHTLPFYTGCTKPTFPMDDTLAVITITFLALVTFIILPG
IRKTRLFWILRVYTSIFIGAVILAVNFSSEWSVGHVANTTYKAFSPKVSVDVGLQ
IGLGVNITFTGTPTVOQLNETINNEAFARLGRSYAEYAKALEKGLPDLVLAEX
FTPRSPCGLVNOYRLAGVYASAMLVAFCLMLANVLSNPLVYGGHMLATGLFOL
LALFFPSMTSLISPCPLRLGTAVLHTHHPAFWITLATGLLCLILGLVMAVAHRMOP
HRLKAFENLSSEDPVLEWSEBEGGLSPHYRSIAESPETQDTPMSVASSETCFKEHP
KESDCSL"
ORIGIN
Query Match 24.2%; Score 348.8; DB 10; Length 1521;
Best Local Similarity 63.5%; Pred. No. 3.2e-63;
Matches 549; Conservative 0; Mismatches 312; Indels 3; Gaps 1;
QY 279 ACTGCTTTTACCCCGAGCCCGGATGCGGAGGCTTCAGCGTTCCACTGCTCATCT 338
Db 305 ACTGCCCTTCTACACTGGACCAAGCAACTTCCAAATGACACTACTCTGGCTGCAT 364
QY 339 TATCTAGTGTGTTTGGCTTACAGCAAGACTTCTCTCATCTTCCGGGGATCCGTGG 398
Db 365 TATCACCATCTTCTCTGACTGACCTGTCTACCTTCATTAATCTGCTGGCATCCGTG 424
QY 399 CCACTGCGGCTGTTTGGTGTGAGAGTCTTCTGCTGTTCTATAGGCGCAGAAT 458
Db 425 GAAGACGAGGCTGTTCTGCTGCTTCCGGGTGAGCAGCACTTATTCATCGGGCTGTAT 484
QY 459 TGTGGCTGTCACTTCACTGAGATGTTGCTGGGTACGTGAACACCAACATCTTA 518
Db 485 CTTGCTGTGAACCTTCACTGAGTGTCTGTGGCCATGTCAACGCCAACACAGTA 544
QY 519 CAAAGCCTTCAAGCGAGCGCGGCTTACAGCCCGTGTGGTCTGCTCGTGGCCTGAGG 578
Db 545 CAAAGCCTTCAAGTCCCAAGTGGTCACTGAGAGTGTGAGAGTGTGAGAGGCGCTG 604
QY 579 CATTATATTAACCTACAGGAGCCCAAGTGCATGAGTGAAGAGAGCACTTGAACAA 638
Db 605 CGTCAACATCACCTTACAGGAGCCCAAGTGCATGAGTGAAGAGCACTTGAACAA 664
QY 639 CGAGCACTTCACTGCGCTGTAAGAAAGAAATTAAGCCCGGAGTACGCGAAGCACTGA 698
Db 665 TGAGGCACTTGCATGCGCGCTGCGGAGAGAGTATGAGAGAGTATGAGAGCGCTGGA 724
QY 699 GAAGGCGCTGCGGAGCCAGTGTCTTACCTGCGGAGAGTTCACACCGAGTAGCCCTTG 758
Db 725 GAAGGCGCTTCCAGACCCAGTGTATACCTGCGTGAAGAGTTCACCCCTGAGAGCCCATG 784
QY 759 CGGCTGTACCAACAGTACCACTGCGGAGAGCACTAGCGCTGCGGCAAGCTATGGGTGC 818
Db 785 TGGCTGTACCAACAGTACCGGCTGCGGAGAGCACTATGCTTGCATGCTGTGGGTGC 844
QY 819 GTTCTGCTTGTGCTCTCTCAACGCTGCTCTCCACGCGGCGCCCGCTCTACGAGG 878
Db 845 ATTCTGCTGCTGCTGCTGCGCAATGTGATGTTGTGATGCGAGTGTGCTGTTATGTGG 904
QY 879 CCTGGCACTGTGACCAACGAGGCTTGGCGCTTTCGGGGCTTTCGCTTGGCC--TC 935
Db 905 CCACATGCTGTGCGCACTGTGCTTTCACAGCTGTGCGCCCTGCTTCTTCTCCATGAC 964
QY 936 CATCTTACGCTGCGGCTTGTGCGGCTTCCAGGCTCTCTCCGCGGCTCACCACTAGTA 995
Db 965 CACGTACTCATATGCGGCTGTCCCTGCGCTTGGGCACTGCTGTGCACTACCA 1024
QY 996 CGGCGCGGCTTGTGGGTACAGCTGCGCAACGCGGCTGCTTCTTCTGGAGGAGGCG 1055
Db 1025 TGGGCGGCTTGTGATCACGTTACGACAGAGTGTGTATTTCTGCTGGCGCTGT 1084
QY 1056 CGTGTGATGCTCCAGTATGTTGCGGCCAGCGCTCTTGCACCTTCTGACCAAGCGC 1115

Db	1085	CATGGCAGTGGCCCA	CAGATGCAGCCTC	ACAGACTGAAGGCTT	CTTCAACCTGAGTTC	1144
QY	1116	CAAGACTGCAGCCAG	GAGAGAG	1139		
Db	1145	AGAGAACCCAGTGTG	AGTGGG	1168		

Search completed: February 23, 2004, 21:57:50
Job time : 5691.97 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 19:33:52 ; Search time 582.643 Seconds

(without alignments)
10506.704 Million cell updates/sec

Title: US-09-936-456-1

Perfect score: 1441

Sequence: 1 aaagtaacgctacagacag.....aaaaaaaaaaaaaaaa 1441

Scoring table: IDENTITY_NUC

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1441	100.0	1441	3	AAA94623 Human CAS
2	1436	99.7	1491	9	ADD19228 Human CDN
3	1435.6	99.6	1474	6	ABL90535 Human pol
4	1413	98.1	1421	3	AAA96505 CDNA enco
5	1392.8	96.7	1460	3	AAC95564 Human sec
6	1392.8	96.7	1460	7	ABZ67235 Human sec
7	1392.8	96.7	1460	7	ABZ73640 Secreted
8	1392.8	96.7	1460	9	ADC20289 Human sec
9	465.4	32.3	498	3	AAA94624 Human CAS
10	405.4	28.1	5033	7	ABZ68115 Human sec
11	405.4	28.1	5033	7	ABZ74587 Secreted
12	405.4	28.1	5033	9	ADC21005 Human sec
13	405	28.1	580	5	ADD19187 Human CDN
14	401.8	27.9	406	5	AA566563 DNA enco
15	396	27.5	2684	7	ADA53627 Human cod
16	356.4	24.7	1636	6	ABNS9923 Novel hum
17	354.8	24.6	1029	5	AAF45134 Human INT
18	354.8	24.6	1029	7	ACD66787 Secreted
19	354.8	24.6	2133	5	AAF45133 Human INT
20	354.8	24.6	2133	7	ABX94107 CDNA enco
21	354.8	24.6	2133	7	ACD66786 Secreted
22	245.8	17.1	2684	7	ADA53627 Human cod
23	161.8	11.2	1338	5	AA567707 DNA enco

24	65.6	4.6	1602	4	ABL21213	Ab121213 Drosophi1
25	65.6	4.6	1808	4	ABL07775	Ab107775 Drosophi1
26	64	4.4	512	4	AAH36185	Aah36185 Human col
27	60.6	4.2	2000	7	ADA71938	Ada71938 Rice gene
28	52.6	3.7	134	7	ABZ68116	Abz68116 Human sec
29	52.6	3.7	134	7	ABZ74588	Abz74588 Secreted
30	52.6	3.7	134	9	ADC21006	Adc21006 Human sec
31	52	3.6	2000	7	ADA71938	Ada71938 Rice gene
32	49	3.4	496	5	ABV56664	Abv56664 Human pro
33	48.6	3.4	1400	4	AA501017	Aa501017 Maize dis
34	47.8	3.3	396	3	AAZ34971	Aaz34971 Wheat bet
35	47.6	3.3	277	7	ABX47508	Abx47508 Human ES
36	47.4	3.3	915	3	AAC60068	Aac60068 Human sec
37	46.8	3.2	2484	5	AA578178	Aa578178 DNA enco
38	46.8	3.2	2589	7	ABZ75977	Abz75977 Human G p
39	46.8	3.2	14055	7	AA161170	Aa161170 Actinosyn
40	46.8	3.2	82746	7	AA161224	Aa161224 Actinosyn
41	46.4	3.2	2206	4	ABK43595	Abk43595 DNA enco
42	46.2	3.2	441	3	AAC69600	Aac69600 Human sec
43	46.2	3.2	531	8	ADA49299	Ada49299 Maize gen
44	46.2	3.2	609	8	ADA49313	Ada49313 Maize gen
45	46.2	3.2				

ALIGNMENTS

RESULT 1
ID AAA94623 standard; DNA; 1441 BP.
AC AAA94623;
DT 11-JAN-2001 (first entry)
DE Human CASB618 coding sequence.
KW Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian;
KW colon; autoimmune disease; HLA_A0201; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 259..1221
FT /*tag= a
FT /product= "Human CASB618"
XX
PN WO200053748-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-EP002048.
XX
PR 11-MAR-1999; 99GB-00005607.
PR 01-SEP-1999; 99GB-00020590.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Bruck CEM, Cascart J, Coche T, Vinals Y De BassolsC;
XX
XX WPI; 2000-572268/53.
XX P-PSDB; AAB26325.
XX
XX New human CASB618 polypeptide, useful as a vaccine for prophylactic and
XX therapeutic treatment of cancers, particularly ovarian or colon cancer,
XX autoimmune diseases and related conditions.
XX
XX Claim 13; Page 61; 76pp; English.
XX
XX The present sequence is the coding sequence of human CASB618 protein. The
XX gene for human CASB618 is thought to be located on chromosome 15. The
XX protein encoded by the present sequence and epitopes of the CASB618
XX protein (see AAB26327 to AAB26399) are useful in diagnosing the occurrence

CC of tumour cells and in vaccines for prophylactic and therapeutic
 CC treatment of cancers, particularly ovarian or colon cancer, autoimmune
 CC diseases and related conditions
 XX

Sequence 1441 BP; 289 A; 466 C; 392 G; 294 T; 0 U; 0 Other;

Query Match 100.0%; Score 1441; DB 3; Length 1441;
 Best Local Similarity 100.0%; Pred. No. 1.3e-311;
 Matches 1441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTAAAGGCTACAGACAGTGAATAATAGTTGCTGCGCGGCTAGAAAACTGTGCG 60
 DB 1 AAAGTAAAGGCTACAGACAGTGAATAATAGTTGCTGCGCGGCTAGAAAACTGTGCG 60
 QY 61 GTACCAACCCAGAGCGGTGAGAGACGCCACCTCCACAGCTTCTTAACGAGAGGTGCA 120
 DB 61 GTACCAACCCAGAGCGGTGAGAGACGCCACCTCCACAGCTTCTTAACGAGAGGTGCA 120
 QY 121 GGAAGTCACTTCAACCAAGCCCACTGCTCCAGCTTGTACGCAAGAAGAGCCCAAGGAC 180
 DB 121 GGAAGTCACTTCAACCAAGCCCACTGCTCCAGCTTGTACGCAAGAAGAGCCCAAGGAC 180
 QY 181 GCGCTCTCCCGCGTCCAGGACAGCCCAAGCTTGTGCTGCTGCGCGCGCTGCGTGCAGC 240
 DB 181 GCGCTCTCCCGCGTCCAGGACAGCCCAAGCTTGTGCTGCTGCGCGCGCTGCGTGCAGC 240
 QY 241 ACTCGCGCGCGTGCAGATGACCCCTGTGGAACGCGGTACTGCTTTTACCCCGAGCCC 300
 DB 241 ACTCGCGCGCGTGCAGATGACCCCTGTGGAACGCGGTACTGCTTTTACCCCGAGCCC 300
 QY 301 CGGCATGCGCGAGGCTTCAAGCGTTCACATGCTCATGCTTATCTAGTGTGCTCTA 360
 DB 301 CGGCATGCGCGAGGCTTCAAGCGTTCACATGCTCATGCTTATCTAGTGTGCTCTA 360
 QY 361 GCAGCAAGCTTCTGCTCATCTTGCAGGAGATCCGTGCGCACTGCGCTGTTGTTG 420
 DB 361 GCAGCAAGCTTCTGCTCATCTTGCAGGAGATCCGTGCGCACTGCGCTGTTGTTG 420
 QY 421 GTGAGAGTCTTCTCACTGCTTCAATAGGCGCAGAAATTGTGCTGTGACATTCAGTGA 480
 DB 421 GTGAGAGTCTTCTCACTGCTTCAATAGGCGCAGAAATTGTGCTGTGACATTCAGTGA 480
 QY 481 GAATGTTGTTGGTGAAGTGAACACCAACACATCTTCAAAAGCTTCAGCGAGCGCGC 540
 DB 481 GAATGTTGTTGGTGAAGTGAACACCAACACATCTTCAAAAGCTTCAGCGAGCGCGC 540
 QY 541 GTTACAGCCCGTGTGCTGCTGCTGCGGCGCTGAGGGCATTAAATTACACTCAGAGG 600
 DB 541 GTTACAGCCCGTGTGCTGCTGCTGCGGCGCTGAGGGCATTAAATTACACTCAGAGG 600
 QY 601 ACCCCAGTGCATCAGCTGGAACGAGACCATTTGACTACACGAGCAGTTCACCTGCGCTG 660
 DB 601 ACCCCAGTGCATCAGCTGGAACGAGACCATTTGACTACACGAGCAGTTCACCTGCGCTG 660
 QY 661 AAAGAGAATTACGCGCGGAGTACGCGAAGCAGTGAAGAGAGGGGCTGCCGAGCCAGTG 720
 DB 661 AAAGAGAATTACGCGCGGAGTACGCGAAGCAGTGAAGAGAGGGGCTGCCGAGCCAGTG 720
 QY 721 CTCTACCTGGCGGAGAAAGTTACACCGAGTAGCCCTTGGCGCTGTACCAACAGTAGCAC 780
 DB 721 CTCTACCTGGCGGAGAAAGTTACACCGAGTAGCCCTTGGCGCTGTACCAACAGTAGCAC 780
 QY 781 CTGGCGGAGACATACGCTCGGCCACGCTATGGTGGCGTCTGCTTCTGCTCTCTCC 840
 DB 781 CTGGCGGAGACATACGCTCGGCCACGCTATGGTGGCGTCTGCTTCTGCTCTCTCC 840
 QY 841 AAGTGTGCTCTCCACGCGCGCGCTCTACGAGAGCCTGCGACTGTGACCAACCGGA 900
 DB 841 AAGTGTGCTCTCCACGCGCGCGCTCTACGAGAGCCTGCGACTGTGACCAACCGGA 900
 QY 901 GCCTTGCAGCTCTTGGGGCTTCTGCGCTTGCCTTCACTCTAGCGTGGCGCTGCGCG 960
 DB 901 GCCTTGCAGCTCTTGGGGCTTCTGCGCTTGCCTTCACTCTAGCGTGGCGCTGCGCG 960

QY 961 CTCGCGCTAAGGCTCTTCCGCGCTACCACTAGTACGGCGCGGCTTCTGGGTACGCTG 1020
 DB 961 CTCGCGCTAAGGCTCTTCCGCGCTACCACTAGTACGGCGCGGCTTCTGGGTACGCTG 1020
 QY 1021 GCAACCGGCGCTCTGCTGCTCTTCTCTGAGAGGGCGGTGTAGTCTCCAGTATGTTCCG 1080
 DB 1021 GCAACCGGCGCTCTGCTGCTCTTCTCTGAGAGGGCGGTGTAGTCTCCAGTATGTTCCG 1080
 QY 1081 CCCAGCGCTCTTGCACCTTCTTGTGAACCAAGCGCCAAAGACTGACAGCAGAGAGAGG 1140
 DB 1081 CCCAGCGCTCTTGCACCTTCTTGTGAACCAAGCGCCAAAGACTGACAGCAGAGAGAGG 1140
 QY 1141 GAGTCACTCTTATCTCGGCGACCCACTGACAGAGCGCGCTCTCCAGACTTAA 1200
 DB 1141 GAGTCACTCTTATCTCGGCGACCCACTGACAGAGCGCGCTCTCCAGACTTAA 1200
 QY 1201 TGTATCACCATAACCTGTGAGGGGACCCCAATCTGGAATCTTCCCGCTTGGGACAT 1260
 DB 1201 TGTATCACCATAACCTGTGAGGGGACCCCAATCTGGAATCTTCCCGCTTGGGACAT 1260
 QY 1261 CGCAGCGCGGGAAGCAGTGCAGCGCCGCGGAGGCTGCGGCGAGAGAGCTCCAGAGGCACTG 1320
 DB 1261 CGCAGCGCGGGAAGCAGTGCAGCGCCGCGGAGGCTGCGGCGAGAGAGCTCCAGAGGCACTG 1320
 QY 1321 AGCGCTGTGCGCGGAGGCTTGGACATCCGACAGGCAACAGGAAAGTCTCTGCGGCGA 1380
 DB 1321 AGCGCTGTGCGCGGAGGCTTGGACATCCGACAGGCAACAGGAAAGTCTCTGCGGCGA 1380
 QY 1381 TCTGTAATTAACCTTTTCTTCTTGTGTTTAA 1440
 DB 1381 TCTGTAATTAACCTTTTCTTCTTGTGTTTAA 1440
 QY 1441 A 1441
 DB 1441 A 1441

RESULT 2

ADD19228 standard; cDNA; 1491 BP.

ADD19228; 15-JAN-2004 (first entry)

Human cDNA from secreted protein gene 45.

human secreted protein; cytosolic; antibacterial; virucide;
 neuroprotective; gynaecological; gastrointestinal-Gen; cardiac;
 cardiovascular-Gen; nephrotropic; antiinflammatory; muscular-Gen;
 respiratory-Gen; immunosuppressive; cerebroprotective; vasotropic;
 neotropic; antiallergic; cancer; bacterial infection; viral infection;
 neural disorder; immune system disorder; blood disorder;
 muscular disorder; reproductive disorder; gastrointestinal disorder;
 pulmonary disorder; cardiovascular disorder; renal disorder;
 inflammatory disorder; proliferative disorder; human; se; gene.

Homo sapiens.

WO2003052377-A2.

26-JUN-2003.

06-NOV-2002; 2002WO-US035606.

07-NOV-2001; 2001US-0331046P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM;

WPI; 2003-533050/50.

DR P-PSDB; ADD19303.

XX

PT New isolated nucleic acids encoding signal transduction pathway component
PT polypeptides, useful for diagnosing, treating, and/or preventing
PT disorders, such as cancer, infections, cardiovascular and inflammatory
PT diseases.

XX

PS Claim 1; SEQ ID NO 55; 554bp; English.

XX

CC The invention relates to an isolated nucleic acid molecule (cDNA)
CC encoding a human secreted protein, representing one of 85 novel genes.

CC

CC Also included are recombinant vectors, host cells (expressing the
CC protein), the secreted proteins (including their fragments, epitopes and
CC homologues), an isolated antibody that binds specifically to the protein,
CC diagnosing a pathological condition or susceptibility to a pathological
CC condition (comprising determining the presence or absence of a mutation
CC in the nucleic acid and diagnosing a condition based on the presence or
CC absence of the mutation), diagnosing a pathological condition or
CC susceptibility to a pathological condition (comprising determining the
CC presence or amount of expression of the protein in a biological sample
CC and diagnosing a condition based on the presence or amount of expression
CC of the protein), preventing, treating or ameliorating a medical condition
CC by administering the nucleic acid or protein to a mammalian subject,
CC identifying a binding partner to the protein, the gene corresponding to
CC the cDNA sequence, and identifying an activity in a biological assay
CC (comprising expressing the nucleic acid in a cell, isolating the
CC supernatant, detecting an activity in a biological assay and identifying
CC the protein in the supernatant having the activity). The nucleic acids
CC and proteins display the following activities: Cytostatic, antibacterial,
CC Virocide, Neuroprotective, Gynaecological, Gastrointestinal-Gen,
CC Cardiant, Cardiovascular-Gen, Nephrotropic, Antiinflammatory, Muscular-
CC Gen, Respiratory-Gen, Immunosuppressive, Cerebroprotective, Vasotropic,
CC Nootropic, Antiallergic. The methods and compositions of the present
CC invention are useful for diagnosing, treating, preventing and/or
CC prognosticating disorders related to the novel polypeptides, such as
CC cancer, bacterial or viral infections, and neural, immune system, blood,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal, inflammatory or proliferative disorders (many examples of these
CC diseases and disorders are given in the specification). The present
CC sequence encodes a novel secreted protein of the invention.

CC

XX Sequence 1491 BP; 293 A; 485 C; 410 G; 302 T; 0 U; 1 Other;

SQ

Query Match 99.7%; Score 1436; DB 9; Length 1491;
Best Local Similarity 99.9%; Pred. No. 1.7e-310;
Matches 1436; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY

1 AAAGTAACGGCTACAGACAGTAGAATAAGTTGCTGCGCGCTAGAAAACTCTGTG 60

Db

44 AAAGTAACGGCTACAGACAGTAGAATAAGTTGCTGCGCGCTAGAAAACTCTGTG 103

QY

61 GTACCAACCCAGAGCGTTGAGAGCAGCCCACTTCCAGCTTCTTAACGAGAGGTGCA 120

Db

104 GTACCAACCCAGAGCGTTGAGAGCAGCCCACTTCCAGCTTCTTAACGAGAGGTGCA 163

QY

121 GGACTCAGACTTCACCAAGCCCACTCGTCCAGCCTTGTACGCAAAAGAGACGCCAAGAC 180

Db

164 GGACTCAGACTTCACCAAGCCCACTCGTCCAGCCTTGTACGCAAAAGAGACGCCAAGAC 223

QY

181 GCGCTCTCCCGCGTCCAGAGCAGCCCAAGCTTGTGCTTGCCTGCCCGCTGCTGACG 240

Db

224 GCGCTCTCCCGCGTCCAGAGCAGCCCAAGCTTGTGCTTGCCTGCCCGCTGCTGACG 283

QY

241 ACTGGCGCGCGTGCAGATGACCCCTGTGGAACGGGCTACTGCTTTTACCCCAAGCC 300

Db

284 ACTGGCGCGCGTGCAGATGACCCCTGTGGAACGGGCTACTGCTTTTACCCCAAGCC 343

QY

301 CGGCATGCCGAGGCTTACGCTTCACTGCTCATGTTATTAGTGTGTTGGCTCTA 360

Db

344 CGGCATGCCGAGGCTTACGCTTCACTGCTCATGTTATTAGTGTGTTGGCTCTA 403

QY

361 GCAGCAAGCTTCCTGCTCATCTTGCGGGGATCCGTGGCCACTCGCGCTGTTTGGTTG 420

Db

404 GCAGCAAGCTTCCTGCTCATCTTGCGGGGATCCGTGGCCACTCGCGCTGTTTGGTTG 463

QY

421 GTGAGAGTTCTTCTCAAGTCTGTTCATATAGCCGAGAAATTGGCTGTGCACTTCACTGCA 480

Db

464 GTGAGAGTTCTTCTCAAGTCTGTTCATATAGCCGAGAAATTGGCTGTGCACTTCACTGCA 523

QY

481 GAATGTTGTTGGGTACAGTGAACACCAACATCTCTACAAAGCCTTACGCCGAGGCGC 540

Db

524 GAATGTTGTTGGGTACAGTGAACACCAACATCTCTACAAAGCCTTACGCCGAGGCGC 583

QY

541 GTTACAGCCCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600

Db

584 GTTACAGCCCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 643

QY

601 ACCCCAGTGCATCAGCTGAACGAGACCAATTGACTACAAAGAGAGAGAGAGAGAGAGAGAG 660

Db

644 ACCCCAGTGCATCAGCTGAACGAGACCAATTGACTACAAAGAGAGAGAGAGAGAGAGAGAG 703

QY

661 AAAGAAATTACGCGCGAGTACGCGAAGCAGCTGAGAAAGGGGCTGCCGAGCCAGTG 720

Db

704 AAAGAAATTACGCGCGAGTACGCGAAGCAGCTGAGAAAGGGGCTGCCGAGCCAGTG 763

QY

721 CTCTACCTGCGGAGAAAGTTCACACCGAGTAGCCCTTGGCGCTGTACCAACAGTACAC 780

Db

764 CTCTACCTGCGGAGAAAGTTCACACCGAGTAGCCCTTGGCGCTGTACCAACAGTACAC 823

QY

781 CTGGCGGACACTACGCTTCGCGCAGCTATGAGTGGCGTTCCTTGTGCTCTCTCC 840

Db

824 CTGGCGGACACTACGCTTCGCGCAGCTATGAGTGGCGTTCCTTGTGCTCTCTCC 883

QY

841 AACGTGCTGCTCTCCACGCGCGCGCGCTCTACGAGAGGCTGCACTGTGACCAACCGGA 900

Db

884 AACGTGCTGCTCTCCACGCGCGCGCGCTCTACGAGAGGCTGCACTGTGACCAACCGGA 943

QY

901 GCCTTGGCGCTCTTGGGCTCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 960

Db

944 GCCTTGGCGCTCTTGGGCTCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 1003

QY

961 CTCGCGCTAGGCTCTCTCGCGCTCACTCACTAGTACGCGCGCGCTTCTGCTGCTGCTG 1020

Db

1004 CTCGCGCTAGGCTCTCTCGCGCTCACTCACTAGTACGCGCGCGCTTCTGCTGCTGCTG 1063

QY

1021 GCAACCGGCGTCTGTGCT 1080

Db

1064 GCAACCGGCGTCTGTGCT 1123

QY

1081 CCCAGCGCTCTTGGCACTCTTGGCACTTGGCACTTGGCACTTGGCACTTGGCACTTGGCACT 1140

Db

1124 CCCAGCGCTCTTGGCACTCTTGGCACTTGGCACTTGGCACTTGGCACTTGGCACTTGGCACT 1183

QY

1141 GGCTACCTCTTATCTCGCGCAGCCCACTGCAAAAGAGCGCGCTCTCCAGACTTAA 1200

Db

1184 GGCTACCTCTTATCTCGCGCAGCCCACTGCAAAAGAGCGCGCTCTCCAGACTTAA 1243

QY

1201 TGTATCACTAATCTGATGAGGAGGAGCCATCTGACTCTTCCCGCTTGGGACAT 1260

Db

1244 TGTATCACTAATCTGATGAGGAGGAGCCATCTGACTCTTCCCGCTTGGGACAT 1303

QY

1261 CGCAGCGCGGAGAGAGTGGCGCGCGAGCCTTGGCGCGAGAGAGCTCCAGAGGCACTG 1320

Db

1304 CGCAGCGCGGAGAGAGTGGCGCGCGAGCCTTGGCGCGAGAGAGCTCCAGAGGCACTG 1363

QY

1321 AGCGTGTGCGCGAGGCTCTGGAATCCGAGGCACTCCAGGAAAGTCTCTGGGGCGA 1380

Db

1364 AGCGTGTGCGCGAGGCTCTGGAATCCGAGGCACTCCAGGAAAGTCTCTGGGGCGA 1423

QY

1381 TCTGTAAATAAACCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1438

Db

1424 TCTGTAAATAAACCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1481

RESULT 3
ABL90535

ID ABL90535 standard; cDNA; 1474 BP.
 XX
 AC ABL90535;
 XX
 DT 24-MAY-2002 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 1097.
 XX
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 OS anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; anticancer;
 PN vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;
 XX cardiatic; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein; gene; ss.
 OS Homo sapiens.
 XX
 PN WO200190304-A2.
 XX
 PD 29-NOV-2001.
 XX
 PF 18-MAY-2001; 2001WO-US016450.
 XX
 PR 19-MAY-2000; 2000US-0205515P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Birse CE, Rosen CA;
 XX
 DR WPI; 2002-122018/16.
 XX
 DR P-PSDB; ABB90126.
 PT
 PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders.
 XX
 PS Claim 4; SEQ ID NO 1097; 2081pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
 CC (ABB8940-ABB9044) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at http://wipo.int/pub/published_pct_sequences
 XX

Query Match	99.6%;	Score 1435.6;	DB 6;	Length 1474;
Best Local Similarity	99.8%;	Pred. No. 2.1e-310;		
Matches 1435; Conservative	2;	Mismatches	1;	Indels 0; Gaps 0

QY	1	AAAGTAACGGGCTACAGACAGTGAGAAATAGTTTCGCTCGCCGGCTAGAAAAA	CTGTGCG	60
Db	27	AAAGTAACGGGCTACAGACAGTGAGAAATAGTTTCGCTCGCCGGCTAGAAAAA	CTGTGCG	86
QY	61	GTACCAACCCCGAGCGCTTGAGAGACAGCCACCTCCACGCTTCTTAACGGAGAG	GTGCA	120
Db	87	GTACCAACCCCGAGCGCTTGAGAGACAGCCACCTCCACGCTTCTTAACGGAGAG	GTGCA	146
QY	121	GGACTCAGACTTCACACAGCCCACTCGTCCCAAGCCTTGATACCAAAAGAGAC	GC	180
Db	147	GGACTCAGACTTCACACAGCCCACTCGTCCCAAGCCTTGATACCAAAAGAGAC	GC	206

QY	181	GGCTCTCCCGCGTCCAGGCAGCCCAAGTGTGCTGTGCTTGCCCTGCGCCGCTGCGTGAC	240
Db	207	GGGCTCTCCCGCGTCCAGGCAGCCCAAGTGTGCTGTGCTTGCCCTGCGCCGCTGCGTGAC	266
QY	241	ACTCGGCCGCGGTGACAGATGACCCCTGTGGAAACGCGCTACTGCTTTTAAACCCCAAGCC	300
Db	267	ACTCGGCCGCGGTGACAGATGACCCCTGTGGAAACGCGCTACTGCTTTTAAACCCCAAGCC	326
QY	301	CGGCATGCGCGAGCTTCAAGCGTTCACCTGCTCATCGTTATTCTAGTGTTTTGGCTCTA	360
Db	327	CGGCATGCGCGAGCTTCAAGCGTTCACCTGCTCATCGTTATTCTAGTGTTTTGGCTCTA	386
QY	361	GCAGCAAGCTTCTGCTCATCTTGCCGCGGATCCGTGGCACTCGCGCTGTTTGGTTG	420
Db	387	GCAGCAAGCTTCTGCTCATCTTRCCGCGGATCCGTGGCACTCGCGCTGTTTGGTTG	446
QY	421	GTGAGAGTTCTTCTCAGTCTGTTCATATGAGCGCAGAAAATTGTGCTGTGCACTTCA	480
Db	447	GTGAGAGTTCTTCTCAGTCTGTTCATATGAGCGCAGAAAATTGTGCTGTGCACTTCA	506
QY	481	GAATGTTCTGGGTACAGTGAACACCAACATCTCAAAAGCCTTCAGCGCAGCGCGC	540
Db	507	GAATGTTCTGGGTACAGTGAACACCAACATCTCAAAAGCCTTCAGCGCAGCGCGC	566
QY	541	GTTACAGCCCGTGTGCGTCTGCTCGTGGGCTTGAGGGCATTAATATTACCTCACAGG	600
Db	567	GTTACAGCCCGTGTGCGTCTGCTCGTGGGCTTGAGGGCATTAATATTACCTCACAGG	626
QY	601	ACCCAGTGCATCAGCTGAACGAGACCATTTGACTACAAGCAGCAGTTCACTGGCGT	660
Db	627	ACCCAGTGCATCAGCTGAACGAGACCATTTGACTACAAGCAGCAGTTCACTGGCGT	686
QY	661	AAAGAGATTACGCGCGGAGTACGCGAACGCACCTGGAGAAAGGGCTGCGGACCCAGT	720
Db	687	AAAGAGATTACGCGCGGAGTACGCGAACGCACCTGGAGAAAGGGCTGCGGACCCAGT	746
QY	721	CTTACCTGGCGGAGAAATTACACCGAGTAGCCCTTGCGGCTGTACCAACAGTACAC	780
Db	747	CTTACCTGGCGGAGAAATTACACCGAGTAGCCCTTGCGGCTGTACCAACAGTACAC	806
QY	781	CTGGCGGACACTACGCTTCGCGCACGCTATGGTGGCGTTCTGCTTGGCTCTCTCC	840
Db	807	CTGGCGGACACTACGCTTCGCGCACGCTATGGTGGCGTTCTGCTTGGCTCTCTCC	866
QY	841	AACGTGCTGCTTCCACGCGCGGCCCGCTCTACGAGGCTGGCAGTCTGACCAACCGGA	900
Db	867	AACGTGCTGCTTCCACGCGCGGCCCGCTCTACGAGGCTGGCAGTCTGACCAACCGGA	926
QY	901	GCTTTCGCGCTCTTCGGGGTCTTGCGCTTGCGCTTCATCTTAGCGTGGCTTGCGCG	960
Db	927	GCTTTCGCGCTCTTCGGGGTCTTGCGCTTGCGCTTCATCTTAGCGTGGCTTGCGCG	986
QY	961	CTTCGCGTAGGCTCTTCGCGGCTCAACCACTCAGTACGCGCGCGCTTCTGGGTCA	1020
Db	987	CTTCGCGTAGGCTCTTCGCGGCTCAACCACTCAGTACGCGCGCGCTTCTGGGTCA	1046
QY	1021	GCAACCGCGCTCTGTGCTCTTCTCGGAGGGGCGCGTGTGAGTCTCCAGTATGTTGG	1080
Db	1047	GCAACCGCGCTCTGTGCTCTTCTCGGAGGGGCGCGTGTGAGTCTCCAGTATGTTGG	1106
QY	1081	CCCAGCGCTTCGCAACCTTCTGACAAAGCGCAAGGACTGCAGCGAGAGAGAGG	1140
Db	1107	CCCAGCGCTTCGCAACCTTCTGACAAAGCGCAAGGACTGCAGCGAGAGAGAGG	1166
QY	1141	GGCTCACCTTTATCTTGCGGCGACCCACTGCAACAAGCAGGCGCTCTCCAGACTTAAA	1200
Db	1167	GGCTCACCTTTATCTTGCGGCGACCCACTGCAACAAGCAGGCGCTCTCCAGACTTAAA	1226
QY	1201	TGTATCACCACTAACCTGTGAGGGGGAACCAATCTGACTCCTTCCCGCCTTGGGACAT	1260
Db	1227	TGTATCACCACTAACCTGTGAGGGGGAACCAATCTGACTCCTTCCCGCCTTGGGACAT	1286

QY	1261	CGCAGGCGCGGGAAGCAGTCCCGCCAGGCGCTGGGCCAGGAGAGCTCCAGGAAGGCACTG	1320
Db	1287	CGCAGGCGCGGGAAGCAGTCCCGCCAGGCGCTGGGCCAGGAGAGCTCCAGGAAGGCACTG	1346
QY	1321	AGCGCTGCTGCGCGGAGGCGCTCGGACATCCGAGGCAACGAGGAAAGTCTCCTGGGGCGA	1380
Db	1347	AGCGCTGCTGCGCGGAGGCGCTCGGACATCCGAGGCAACGAGGAAAGTCTCCTGGGGCGA	1406
QY	1381	TCTGTAAATAAACCTTTTTTCTTTTGTTTTTTAAAAAATAAAAAAAAAAAAAA	1438
Db	1407	TCTGTAAATAAACCTTTTTTCTTTTGTTTTTTAAAAAATAAAAAAAAAAAAAA	1464

RESULT 4	
AAA96505	
ID	AAA96505 standard; cDNA; 1421 BP.

AC AAA965057

DT 08-FEB-2001 (first entry)

DE cDNA encoding a human transmembrane protein.

KW Human; transmembrane protein; cell proliferation disorder; myeloma;
KW reproductive disorder; smooth muscle disorder; neurological disorder;
KW arteriosclerosis; leukemia; acquired immunodeficiency syndrome; AIDS;
KW allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;
KW Alzheimer's disease; Tourette's disorder; ss.

05 Homo sapiens.

	Key	Location/Qualifiers
EH		267. .1229
FT	CDS	/*tag= a
FT		

PN WO2000056891-A2

PD 28-SEP-2000.

PF 22-MAR-2000; 2000WO-US007817.

PR 22-MAR-1999; 99US-0125537P.

⊗

1. **Introduction**
 2. **Background**
 3. **Methodology**
 4. **Results**
 5. **Discussion**
 6. **Conclusion**
 7. **References**
 8. **Appendix**
 9. **Index**
 10. **Table of Contents**
 11. **Figure 1**
 12. **Figure 2**
 13. **Figure 3**
 14. **Figure 4**
 15. **Figure 5**
 16. **Figure 6**
 17. **Figure 7**
 18. **Figure 8**
 19. **Figure 9**
 20. **Figure 10**
 21. **Figure 11**
 22. **Figure 12**
 23. **Figure 13**
 24. **Figure 14**
 25. **Figure 15**
 26. **Figure 16**
 27. **Figure 17**
 28. **Figure 18**
 29. **Figure 19**
 30. **Figure 20**
 31. **Figure 21**
 32. **Figure 22**
 33. **Figure 23**
 34. **Figure 24**
 35. **Figure 25**
 36. **Figure 26**
 37. **Figure 27**
 38. **Figure 28**
 39. **Figure 29**
 40. **Figure 30**
 41. **Figure 31**
 42. **Figure 32**
 43. **Figure 33**
 44. **Figure 34**
 45. **Figure 35**
 46. **Figure 36**
 47. **Figure 37**
 48. **Figure 38**
 49. **Figure 39**
 50. **Figure 40**
 51. **Figure 41**
 52. **Figure 42**
 53. **Figure 43**
 54. **Figure 44**
 55. **Figure 45**
 56. **Figure 46**
 57. **Figure 47**
 58. **Figure 48**
 59. **Figure 49**
 60. **Figure 50**
 61. **Figure 51**
 62. **Figure 52**
 63. **Figure 53**
 64. **Figure 54**
 65. **Figure 55**
 66. **Figure 56**
 67. **Figure 57**
 68. **Figure 58**
 69. **Figure 59**
 70. **Figure 60**
 71. **Figure 61**
 72. **Figure 62**
 73. **Figure 63**
 74. **Figure 64**
 75. **Figure 65**
 76. **Figure 66**
 77. **Figure 67**
 78. **Figure 68**
 79. **Figure 69**
 80. **Figure 70**
 81. **Figure 71**
 82. **Figure 72**
 83. **Figure 73**
 84. **Figure 74**
 85. **Figure 75**
 86. **Figure 76**
 87. **Figure 77**
 88. **Figure 78**
 89. **Figure 79**
 90. **Figure 80**
 91. **Figure 81**
 92. **Figure 82**
 93. **Figure 83**
 94. **Figure 84**
 95. **Figure 85**
 96. **Figure 86**
 97. **Figure 87**
 98. **Figure 88**
 99. **Figure 89**
 100. **Figure 90**
 101. **Figure 91**
 102. **Figure 92**
 103. **Figure 93**
 104. **Figure 94**
 105. **Figure 95**
 106. **Figure 96**
 107. **Figure 97**
 108. **Figure 98**
 109. **Figure 99**
 110. **Figure 100**
 111. **Figure 101**
 112. **Figure 102**
 113. **Figure 103**
 114. **Figure 104**
 115. **Figure 105**
 116. **Figure 106**
 117. **Figure 107**
 118. **Figure 108**
 119. **Figure 109**
 120. **Figure 110**
 121. **Figure 111**
 122. **Figure 112**
 123. **Figure 113**
 124. **Figure 114**
 125. **Figure 115**
 126. **Figure 116**
 127. **Figure 117**
 128. **Figure 118**
 129. **Figure 119**
 130. **Figure 120**
 131. **Figure 121**
 132. **Figure 122**
 133. **Figure 123**
 134. **Figure 124**
 135. **Figure 125**
 136. **Figure 126**
 137. **Figure 127**
 138. **Figure 128**
 139. **Figure 129**
 140. **Figure 130**
 141. **Figure 131**
 142. **Figure 132**
 143. **Figure 133**
 144. **Figure 134**
 145. **Figure 135**
 146. **Figure 136**
 147. **Figure 137**
 148. **Figure 138**
 149. **Figure 139**
 150. **Figure 140**
 151. **Figure 141**
 152. **Figure 142**
 153. **Figure 143**
 154. **Figure 144**
 155. **Figure 145**
 156. **Figure 146**
 157. **Figure 147**
 158. **Figure 148**
 159. **Figure 149**
 160. **Figure 150**
 161. **Figure 151**
 162. **Figure 152**
 163. **Figure 153**
 164. **Figure 154**
 165. **Figure 155**
 166. **Figure 156**
 167. **Figure 157**
 168. **Figure 158**
 169. **Figure 159**
 170. **Figure 160**
 171. **Figure 161**
 172. **Figure 162**
 173. **Figure 163**
 174. **Figure 164**
 175. **Figure 165**
 176. **Figure 166**
 177. **Figure 167**
 178. **Figure 168**
 179. **Figure 169**
 180. **Figure 170**
 181. **Figure 171**
 182. **Figure 172**
 183. **Figure 173**
 184. **Figure 174**
 185. **Figure 175**
 186. **Figure 176**
 187. **Figure 177**
 188. **Figure 178**
 189. **Figure 179**
 190. **Figure 180**
 191. **Figure 181**
 192. **Figure 182**
 193. **Figure 183**
 194. **Figure 184**
 195. **Figure 185**
 196. **Figure 186**
 197. **Figure 187**
 198. **Figure 188**
 199. **Figure 189**
 200. **Figure 190**
 201. **Figure 191**
 202. **Figure 192**
 203. **Figure 193**
 204. **Figure 194**
 205. **Figure 195**
 206. **Figure 196**
 207. **Figure 197**
 208. **Figure 198**
 209. **Figure 199**
 210. **Figure 200**
 211. **Figure 201**
 212. **Figure 202**
 213. **Figure 203**
 214. **Figure 204**
 215. **Figure 205**
 216. **Figure 206**
 217. **Figure 207**
 218

PI Baughn MR, Lu DAM, Azimzai Y, Yang J;

WPI; 2000-579485/54.

PT New human transmembrane proteins are used to treat a disease or condition
PT associated with decreased expression of functional HTMP e.g. Tourette's
PT disorder, angina and leukemia.

PS Claim 4; Page 129; 130pp; English.

CC The present sequence encodes a human transmembrane proteins (HTMP).
CC Agonists and antagonists of the protein are used to treat a disease or
CC condition associated with overexpression of the protein. Diseases and
CC conditions which can be treated include cell proliferative,
CC immunological, reproductive, smooth muscle and neurological disorders
CC e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency
CC syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,
CC stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The
CC polynucleotides may be used to detect and quantify gene expression in
CC biopsied tissues where protein expression may be correlated with disease
CC e.g. to determine absence, presence or excess expression of HTMP or to
CC monitor regulation of HTMP expression during therapeutic intervention
XX
SQ Sequence 1421 BP; 264 A; 468 C; 395 G; 294 T; 0 U; 0 Other;

Query Match	98.1%;	Score 1413;	DB 3;	Length 1421;
Best Local Similarity	100.0%;	Pred. NO. 2.3e-305;		
Matches 1413; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	AAAGTAA	CGGCTTACAGACAGT	GAGAAATAGTTTGGCTCGCCGGCTAGAAAAACTCTGTG	60	
Db	9	AAAGTAA	CGGCTTACAGACAGT	GAGAAATAGTTTGGCTCGCCGGCTAGAAAAACTCTGTG	68	
QY	61	GTACCA	ACCCAGAGCGTTG	AGACAGCCCACTCCACGCTTCTTAA	CGAGAGGTGCA 120	
Db	69	GTACCA	ACCCAGAGCGTTG	AGAGCAGCCCACTCCACGCTTCTTAA	CGAGAGGTGCA 128	
QY	121	GGACT	CAGACTTCA	CCAGCCCACTGGTCCCA	GCTTTGTA	CGCAAAGACGCCCAAGAC 180
Db	129	GGACT	CAGACTTCA	CCAGCCCACTGGTCCCA	GCTTTGTA	CGCAAAGACGCCCAAGAC 188
QY	181	GGCGT	TCCCCGCGTCC	AGAGACCCCAAGCTTGTGCTTGCCCTGCCCCCTGCGTAGC	240	
Db	189	GGCGT	TCCCCGCGTCC	AGAGACCCCAAGCTTGTGCTTGCCCTGCCCCCTGCGTAGC	248	
QY	241	ACTCG	CGCGCGGTG	CAGCATGACCCCTGTG	GAACGGCGTACTGCTTTTAA	CCCCCAGCCC 300
Db	249	ACTCG	CGCGCGGTG	CAGCATGACCCCTGTG	GAACGGCGTACTGCTTTTAA	CCCCCAGCCC 308
QY	301	CGGCA	TGCCGAGCGCTT	CAGCGTTCCACTGCTCATCGTATTTCTAGT	TTTTGGCTCTA 360	
Db	309	CGGCA	TGCCGAGCGCTT	CAGCGTTCCACTGCTCATCGTATTTCTAGT	TTTTGGCTCTA 368	
QY	361	GCAGC	AGCTTCTGCTCATCTT	GCAGGGAATCCGTGGCCACTCGCGCTGTTTGGTTG	420	
Db	369	GCAGC	AGCTTCTGCTCATCTT	GCAGGGAATCCGTGGCCACTCGCGCTGTTTGGTTG	428	
QY	421	GTGAG	TCTCTTCTCAGTCTGTT	CATAGCGCAGAAATGTGGCTGTGCA	CTTCACTGCA 480	
Db	429	GTGAG	TCTCTTCTCAGTCTGTT	CATAGCGCAGAAATGTGGCTGTGCA	CTTCACTGCA 488	
QY	481	GAATG	TGTGTGGTACAGTGA	CAACCAACATCTCAAAAGCCTTCA	GCAGCGCGC 540	
Db	489	GAATG	TGTGTGGTACAGTGA	CAACCAACATCTCAAAAGCCTTCA	GCAGCGCGC 548	
QY	541	GTTAC	AGCCCGTGTGCTGTCT	GTGGGCTGGAGGGCATTAATATTA	CACTCACAGGG 600	
Db	549	GTTAC	AGCCCGTGTGCTGTCT	GTGGGCTGGAGGGCATTAATATTA	CACTCACAGGG 608	
QY	601	ACCCAG	TGCATCAGCTGA	ACGACCACTTGA	CTACAACGACAGCATTCACCTGGCGTCTG 660	
Db	609	ACCCAG	TGCATCAGCTGA	ACGACCACTTGA	CTACAACGACAGCATTCACCTGGCGTCTG 668	
QY	661	AAAGAG	AATTACGCGCGGAGTA	CGCGAACCACTG	GAAGAGGGGCTGCCGACCAAGTG 720	
Db	669	AAAGAG	AATTACGCGCGGAGTA	CGCGAACCACTG	GAAGAGGGGCTGCCGACCAAGTG 728	
QY	721	CTTAC	CTGGCGGAGTA	GTTCACACCGAGTA	AGCCCTTGGGGCTGTACCA	CCAGTACAC 780
Db	729	CTTAC	CTGGCGGAGTA	GTTCACACCGAGTA	AGCCCTTGGGGCTGTACCA	CCAGTACAC 788
QY	781	CTGG	CGGACACTACGCTCG	GCACAGCTATGGGTGGCGTTCTGCTT	CTGCTCTCC 840	
Db	789	CTGG	CGGACACTACGCTCG	GCACAGCTATGGGTGGCGTTCTGCTT	CTGCTCTCC 848	
QY	841	AACGT	GTGCTTCCACGCGCGCCGCTCTA	CGAGGCTTGACATGTGACCA	CCGGA 900	
Db	849	AACGT	GTGCTTCCACGCGCGCCGCTCTA	CGAGGCTTGACATGTGACCA	CCGGA 908	
QY	901	GCCTT	CGCGCTCTTGGGGTCTT	CGCCTTGACCTCATCTT	TAGCGTGCCTGCCCCG 960	
Db	909	GCCTT	CGCGCTCTTGGGGTCTT	CGCCTTGACCTCATCTT	TAGCGTGCCTGCCCCG 968	
QY	961	CTCCG	CTTAGGCTCTCCGCGCTAC	CACTCAGTACGCGCGCCTT	CTGGGTCA	CGCTG 1020
Db	969	CTCCG	CTTAGGCTCTCCGCGCTAC	CACTCAGTACGCGCGCCTT	CTGGGTCA	CGCTG 1028
QY	1021	GCAAC	CGCGCTCTGTGCTCTT	CTTCGAGGGGCGGTG	GTAGTCTCCAGTATGTTGG 1080	

Db 1029 GCAACCGCGCTCCTGCTGCTCTTCTCGAGAGGCGCGTGAGTCTCCAGATATGTTCCG 1088
QY 1081 CCCAGGCTCTTTCGACCCCTTCTGACCCAAAGCGCCAAAGACTGACGCCAGAGAGAGG 1140
Db 1089 CCCAGGCTCTTTCGACCCCTTCTGACCCAAAGCGCCAAAGACTGACGCCAGAGAGAGG 1148
QY 1141 GGCTACCTCTTATCTCTGCGGACCCACTGCACAGCAGGCGCTCTCCAGACTTAA 1200
Db 1149 GGCTACCTCTTATCTCTGCGGACCCACTGCACAGCAGGCGCTCTCCAGACTTAA 1208
QY 1201 TGTATACCACTAACCCTGTGAGGGGACCCAACTGTGACTCCTTCCCGCTTGAGACAT 1260
Db 1209 TGTATACCACTAACCCTGTGAGGGGACCCAACTGTGACTCCTTCCCGCTTGAGACAT 1268
QY 1261 CGCAGGCGGGAAGCAGTGCCTCCGACGCTGGGCCAGAGAGACTCCAGAGAGGCACTG 1320
Db 1269 CGCAGGCGGGAAGCAGTGCCTCCGACGCTGGGCCAGAGAGACTCCAGAGAGGCACTG 1328
QY 1321 AGCGCTGCTGGCGGAGGCTCGGACATCCGACAGCACCAGGAAAGTCTCTGGGCGCA 1380
Db 1329 AGCGCTGCTGGCGGAGGCTCGGACATCCGACAGCACCAGGAAAGTCTCTGGGCGCA 1388
QY 1381 TCTGTAATAAACCTTTTCTTTCTTTT 1413
Db 1389 TCTGTAATAAACCTTTTCTTTCTTTT 1421

RESULT 5
AAC95564
ID AAC95564 standard; cDNA; 1460 BP.

AC AAC95564;
DT 21-FEB-2001 (first entry)

DE Human secreted protein gene 44 SEQ ID NO:54.

XX Human; secreted protein; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antitumor; vulnery; anticonvulsant; antibacterial;
KW antifungal; antiparasitic; cardiant; cancer; immune disease; allergy;
KW cardiovascular disorder; wound healing; infection; neurological disease;
ss.

OS Homo sapiens.

PN WO200061596-A1.

PD 19-OCT-2000.

PF 06-APR-2000; 2000WO-US008983.

PR 09-APR-1999; 99US-0128703P.

PR 14-JAN-2000; 2000US-0176068P.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

PI Ruben SM, Komatsoulis G;

DR WPI; 2000-611865/58.

DR P-PSDB; AAB52055.

XX Fifty nucleic acid molecules encoding human secreted proteins, useful in

PT the prevention, treatment and diagnosis of cancer, immune disorders,

PT cardiovascular disorders and neurological diseases.

XX Claim 1; Page 443; 505bp; English.

CC Polynucleotide sequences AAC95521 - AAC95570 represent cDNA encoding 50

CC human secreted proteins AAB52012 - AAB52161. Sequences AAB52062 -

CC AAB52103 represent alternative polypeptides encoded by the genes, and

CC amino acid sequences with which they share homology. The genes and
CC proteins have activities dependent on the tissues and cells in which they
CC are expressed. Examples of their activities include cytostatic;
CC immunosuppressive; nootropic; neuroprotective; antiviral; antiallergic;
CC hepatotropic; antidiabetic; antiinflammatory; antitumor; vulnery;
CC anticonvulsant; antibacterial; antifungal; antiparasitic; and cardiant.
CC The secreted proteins, polypeptides, antagonists and agonists may be
CC useful in treating, preventing and/or diagnosing diseases and disorders
CC such as cancer, particularly breast and ovarian cancer, and other cancers
CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,
CC liver, lung, or urogenital. Immune disorders such as Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases such as viral,
CC bacterial, fungal and parasitic infections may also be treated using the
CC proteins and polynucleotides of the invention. Sequences AAC95512 -
CC AAC95520 and AAB52011 are used in the isolation and characterisation of
CC the proteins and polynucleotides of the invention

XX Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;

Query Match 96.7%; Score 1392.8; DB 3; Length 1460;
Best Local Similarity 99.3%; Pred. No. 7.5e-301;
Matches 1430; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

QY 1 AAAGTAACGCTACAGACAGTGAAGAAATAGTTGCTCGCCGCGCTAGAAAACCTGTGCG 60
Db 22 AAAGTAACGCTACAGACAGTGAAGAAATAGTTGCTCGCCGCGCTAGAAAACCTGTGCG 81
QY 61 GTACCAACCCAGAGCGTTGAGAGCAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 120
Db 82 GTACCAACCCAGAGCGTTGAGAGCAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 141
QY 121 GCACTCAGACTTACCAAGCCCACTCGGTCCAGCGCTTGTACGCAAGAGAGCGCAAGAC 180
Db 142 GCACTCAGACTTACCAAGCCCACTCGGTCCAGCGCTTGTACGCAAGAGAGCGCAAGAC 201
QY 181 GCGCTCTCCGCGTCCAGAGCAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db 202 GCGCTCTCCGCGTCCAGAGCAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 261
QY 241 ACTCGCGCGGTGACAGTACCCCTGTGAACGGGTACTGCTTTTACCCCAAGCCC 300
Db 262 ACTCGCGCGGTGACAGTACCCCTGTGAACGGGTACTGCTTTTACCCCAAGCCC 321
QY 301 CGGCATGCCGAGCGCTTACAGCGTTCCACTGCTCAGCTTATTTAGTGTGTTTGGCTCTA 360
Db 322 CGGCATGCCGAGCGCTTACAGCGTTCCACTGCTCAGCTTATTTAGTGTGTTTGGCTCTA 381
QY 361 GCAGCAAGCTTCTGCTCACTTCTGCGGGAGATCCGTGGCCACTCGGCGCTGTTTGGTTG 420
Db 382 GCAGCAAGCTTCTGCTCACTTCTGCGGGAGATCCGTGGCCACTCGGCGCTGTTTGGTTG 441
QY 421 GTGAGAGTTCTTCTCAGTGTGTATAGCGCGCAAAATTTGGCTGTGCACTTCACTGCA 480
Db 442 GTGAGAGTTCTTCTCAGTGTGTATAGCGCGCAAAATTTGGCTGTGCACTTCACTGCA 501
QY 481 GAATGTTCTGTTGTTACAGTGAACCAACATCTCTCAAAAGCCTTACGGCAGCGCGC 540
Db 502 GAATGTTCTGTTGTTACAGTGAACCAACATCTCTCAAAAGCCTTACGGCAGCGCGC 561
QY 541 GTTACAGCCCGTGTGCTGCTGCTGCGGCTGAGGGCATTAATATTACACTCAGAGG 600
Db 562 GTTACAGCCCGTGTGCTGCTGCTGCGGCTGAGGGCATTAATATTACACTCAGAGG 621
QY 601 ACCCCAGTGCATCAGCTGAAGAGACCAATTGACTACAGAGCAGTTCACCTGGGCTG 660
Db 622 ACCCCAGTGCATCAGCTGAAGAGACCAATTGACTACAGAGCAGTTCACCTGGGCTG 681
QY 661 AAAGAGATTAGCGCGGAGTACGCGAAGCAGCACTGGAAGAGGGCTGCGGACCAAGTG 720

Db 682 AAAGAGATTACGCCGGAGTAGCGGAAACGACTGGAGAAAGGGCTGCGGACCAGTG 741
QY 721 CTCTACCTGGCGAGAGATTACACCGAGTAGCCCTTGCGGCTGTACCAGCAGTACCAC 780
Db 742 CTCTACTGGCGAGAGATTACACCGAGTAGCCCTTGCGGCTGTACCAGCAGTACCAC 801
QY 781 CTGGCGGACACTACGCGCTGCGGACAGCTATGGGTGGCTTCTGCTTCTGCTCTCC 840
Db 802 CTGGCGGACACTACGCGCTGCGGACAGCTATGGGTGGCTTCTGCTTCTGCTCTCC 861
QY 841 AACGTGCTGCTCTCCAGCGCGCGCCGCTCTACGAGGCGCTGGCACTGTGACACCGGA 900
Db 862 AACGTGCTGCTCTCCAGCGCGCGCCGCTCTACGAGGCGCTGGCACTGTGACACCGGA 921
QY 901 GCCTTCGGGCTCTTCGGGGTCTTCGGCTTGGCCCTTCATCTTAAGCGTGCGGCTTGCCCG 960
Db 922 GCCTTCGGGCTCTTCGGGGTCTTCGGCTTGGCCCTTCATCTTAAGCGTGCGGCTTGCCCG 981
QY 961 CTCGCGCTAGGCTCTCCGCGGCTCACCACCTAGTACGGCGCGGCTTCTGGGTACGCTG 1020
Db 982 CTCGCGCTAGGCTCTCCGCGGCTCACCACCTAGTACGGCGCGGCTTCTGGGTACGCTG 1041
QY 1021 GCAACCGGCGTCTGTGCTCTTCTCTGAGGGGCGGCTGTGAGTCTCCAGTATGTTGG 1080
Db 1042 GCAACCGGCGTCTGTGCTCTTCTCTGAGGGGCGGCTGTGAGTCTCCAGTATGTTGG 1101
QY 1081 CCCAGCGCTCTTGCGACCCCTTGTGACCAAGCGCCAAAGAGACTGC-AGCCAGAGAGAG 1139
Db 1102 CCCAGCGCTCTTGCGACCCCTTGTGACCAAGCGCCAAAGAGACTGCAGAGAGAGGG 1161
QY 1140 GGGCTACCTCTTATCTCGGGGACCCCACTGCAAGAGAGCGGCTCTCCAGAC-TTAA 1198
Db 1162 GGGCTACCTCTTATCTCGGGGACCCCACTGCAAGAGAGCGGCTCTCCAGACTTTAA 1220
QY 1199 AATGTATACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTCCCGGCTTGGAG 1258
Db 1221 AATGTATACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTCCCGGCTTGGAG 1280
QY 1259 ATCGAGGCGCGGAGAGAGAGTGCAGGCGGCTGAGGCGGCTGAGAGTCCAGAGAGGG 1318
Db 1281 ATCGAGGCGCGGAGAGAGAGTGCAGGCGGCTGAGGCGGCTGAGAGTCCAGAGAGGG 1340
QY 1319 TGAGCGCTGCTGGCGGAGGCTCGGACATCCGAGGACACGAGAAAGTCTCTGGGCG 1378
Db 1341 TGAGCGCTGCTGGCGGAGGCTCGGACATCCGAGGACACGAGAAAGTCTCTGGGCG 1400
QY 1379 GATCTGTAATAAACCTTTTCTTTTGTGTTTAAAAAAAAAAAAAAAAAAAAAAA 1438
Db 1401 GATCTGTAATAAACCTTTTCTTTTGTGTTTAAAAAAAAAAAAAAAAAAAAAAA 1460

RESULT 6

ABZ67235
ID ABZ67235 standard; cDNA; 1460 BP.
XX
AC ABZ67235;
XX
DT 26-MAR-2003 (first entry)
XX
DE Human secreted protein encoding cDNA SEQ ID NO 355.
XX
KW Human; secreted protein; nootropic; neuroprotective; cytosolic;
KW virucide; dermatological; immunosuppressive; anti-inflammatory; anti-HIV;
KW vulnerary; antibacterial; antiparkinsonian; antispasmodic; antianaemic;
KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
KW antiinflammatory; antiallergic; antidiabetic; antitumor; anticonvulsant;
KW antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;
KW cardiovascular disorder; neurological disease; nephrotropic;
KW gene therapy; gene; chromosome 9p21; ds.
XX
OS Homo sapiens.
XX
PN WO200277186-A2.

XX 03-OCT-2002.
PD
XX
PF 26-MAR-2002; 2002WC-US009188.
XX
PR 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2003-040583/03.
DR P-PSDB; ABP99814.
XX
PT New human secreted proteins encoded by genes contained in cDNA clones
PT (e.g. HGAC19), useful for preventing, treating or diagnosing e.g. AIDS,
PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
PT West Nile fever.
XX
PS Claim 7; Page 1353; 2423pp; English.
XX
CC The invention relates to novel human genes (ABZ66891-ABZ68209) and the
CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections
SQ Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;
Query Match 96.7%; Score 1392.8; DB 7; Length 1460;
Best local Similarity 99.3%; Pred. No. 7.5e-301;
Matches 1430; Conservative 0; Mismatches 7; Indels 3; Gaps 3;
QY 1 AANGTACGGGCTACAGACAGTGAAGAATATGTTGCTGCGCGGCTAGAAAACTCTGCG 60
Db 22 AANGTACGGGCTACAGACAGTGAAGAATATGTTGCTGCGCGGCTAGAAAACTCTGCG 81
QY 61 GTACCAACCCGAGAGCGTTGAGAGAGAGCCACCTCCAGGCTTCTTAACGAGAGGTGCA 120
Db 82 GTACCAACCCGAGAGCGTTGAGAGAGAGCCACCTCCAGGCTTCTTAACGAGAGGTGCA 141
QY 121 GGAAGTACAGCTTACAGAGCCCACTGGTCCAGGCTTGTACGCAAAAGAGAGCCAAAGGAC 180
Db 142 GGAAGTACAGCTTACAGAGCCCACTGGTCCAGGCTTGTACGCAAAAGAGAGCCAAAGGAC 201
QY 181 GCGCTCTCCCGGCTCCAGAGAGCCCACTGGTCCAGGCTTGTACGCAAAAGAGAGCCAAAGGAC 240
Db 202 GCGCTCTCCCGGCTCCAGAGAGCCCACTGGTCCAGGCTTGTACGCAAAAGAGAGCCAAAGGAC 261
QY 241 ACTGCGCGGCGTCCAGAGATGACCTGTGAAACGGCGTAAGCTTTTAAACCCAGAGCC 300
Db 262 ACTGCGCGGCGTCCAGAGATGACCTGTGAAACGGCGTAAGCTTTTAAACCCAGAGCC 321
QY 301 CGGATGCGCGAGGCTTCAAGGTTCACTGCTCATGTTATTTAGTGTGTTTGGCTCTA 360
Db 322 CGGATGCGCGAGGCTTCAAGGTTCACTGCTCATGTTATTTAGTGTGTTTGGCTCTA 381
QY 361 GCAGCAAGCTTCTGCTCATTTGCGGGGATCCGTGGCCACTCGCGCTGTTGGTTG 420
Db 382 GCAGCAAGCTTCTGCTCATTTGCGGGGATCCGTGGCCACTCGCGCTGTTGGTTG 441

QY 421 GTGAGAGTTCTTCTCAGTCTGTTCATAGCGCGAGAAATGTGGCTGTGCACTTCACTGCA 480
DB 442 GTGAGAGTTCTTCTCAGTCTGTTCATAGCGCGAGAAATGTGGCTGTGCACTTCACTGCA 501
QY 481 GAATGGTTCGTGGGTACAGTGAACACCAACACATCTTACAAAGCCTTCAAGCGGAGCGGC 540
DB 502 GAATGGTTCGTGGGTACAGTGAACACCAACACATCTTACAAAGCCTTCAAGCGGAGCGGC 561
QY 541 GTTACAGCCCCGTGTGGTCTGTCTGTGGGCTGGAAGGGCATTAATATACACTACACAGG 600
DB 562 GTTACAGCCCCGTGTGGTCTGTCTGTGGGCTGGAAGGGCATTAATATACACTACACAGG 621
QY 601 ACCCAGTGCATCAGCTGAACGAGACCATTGACTACAACGAGCAGTTCACTGGCGGTG 660
DB 622 ACCCAGTGCATCAGCTGAACGAGACCATTGACTACAACGAGCAGTTCACTGGCGGTG 681
QY 661 AAAGAGATTACCGCCCGGAGTACCGGAAACGCACTGGAGAGGGGCTGCGGAGCCAGTG 720
DB 682 AAAGAGATTACCGCCCGGAGTACCGGAAACGCACTGGAGAGGGGCTGCGGAGCCAGTG 741
QY 721 CTCTACCTGGCGGAGAGTTTACACCGAGTAGCCCTTGGCGCTGTACCAAGTACAC 780
DB 742 CTCTACCTGGCGGAGAGTTTACACCGAGTAGCCCTTGGCGCTGTACCAAGTACAC 801
QY 781 CTGGCGGACACTAGCCTCGGCCACGCTATGGGTGGCGTCTGCTTCTGCTCTCTCC 840
DB 802 CTGGCGGACACTAGCCTCGGCCACGCTATGGGTGGCGTCTGCTTCTGCTCTCTCC 861
QY 841 AACGTGCTCTCTCCAGCGCGCGCTCTACGAGAGGCTGACCTGTGACCAACCGGA 900
DB 862 AACGTGCTCTCTCCAGCGCGCGCTCTACGAGAGGCTGACCTGTGACCAACCGGA 921
QY 901 GCCTTCGCGCTCTTCGCGGCTCTTCGCTTGGCTTCATCTAGCGTGGCTCTGCGCG 960
DB 922 GCCTTCGCGCTCTTCGCGGCTCTTCGCTTGGCTTCATCTAGCGTGGCTCTGCGCG 981
QY 961 CTCGCGCTAGGCTCTCTCGCGCTCACTCAGTACGGCGCGCTTCTGGGTCAAGCTG 1020
DB 982 CTCGCGCTAGGCTCTCTCGCGCTCACTCAGTACGGCGCGCTTCTGGGTCAAGCTG 1041
QY 1021 GCAACCGCGCTCTTCGCTTCTTCGAGGAGGCGCTGTGAGTCTCCAGTATGTTGG 1080
DB 1042 GCAACCGCGCTCTTCGCTTCTTCGAGGAGGCGCTGTGAGTCTCCAGTATGTTGG 1101
QY 1081 CCCAGCGCTCTTCGCACTCTTGAGCAACCAAGCGCCAAAGGACTGC-AGCCAGAGAGAGG 1139
DB 1102 CCCAGCGCTCTTCGCACTCTTGAGCAACCAAGCGCCAAAGGACTGCAGGCCAGAGAGG 1161
QY 1140 GGGCTCACCTCTTATCTTCGCGAGCCCACTGCACAGAGCGCGCTCTCCAGAC-TTAA 1198
DB 1162 GGGCTCACCTCTTATCT- GGGAGCCCACTGCACAGAGCGCGCTTCCAGACTTTAA 1220
QY 1199 AATGTATACCACTAATCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGGAC 1258
DB 1221 AATGTATACCACTAATCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGGAC 1280
QY 1259 ATCGCAGCGCGGAGAGAGTGCCCGCAGGCTGGGCCAGAGAGCTCCAGAGAGGCGAC 1318
DB 1281 ATCGCAGCGCGGAGAGAGTGCCCGCAGGCTGGGCCAGAGAGCTCCAGAGAGGCGAC 1340
QY 1319 TGAGCGCTGTGGCGGAGGCTCGGACATCCGACAGGACACAGGAGAAAGTCTCTGGGCG 1378
DB 1341 TGAGCGCTGTGGCGGAGGCTCGGACATCCGACAGGACACAGGAGAAAGTCTCTGGGCG 1400
QY 1379 GATCTGTAAATAAACCTTTTCTTTGTTTAAAAAATTTTAAAAAATTTTAAAAA 1438
DB 1401 GATCTGTAAATAAACCTTTTCTTTGTTTAAAAAATTTTAAAAAATTTTAAAAA 1460

RESULT 7
ABZ73640
ID ABZ73640 standard; cDNA; 1460 BP.
XX

AC ABZ73640;
XX
DT 12-MAY-2003 (first entry)
XX
DE Secreted protein-encoding gene 360 cDNA clone HUFCL31, SEQ ID NO:370.
XX
KM Human; secreted protein; cancer; tumour; hyperproliferative disorder;
KM autoimmune disorder; inflammation; angiogenic diseases; AIDS;
KM acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
KM drug screening; chromosome identification; chromosome mapping;
KM cytosolic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
KM antianaemic; vulnery; chromosome 9p21; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200277013-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-US009370.
XX
PR 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-040578/03.
DR P-PSDB; ABR01306.
XX
PT New human secreted proteins and nucleic acids, useful for detecting or
PT treating cancer or other hyperproliferative disorders, autoimmune
PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
XX
PS Claim 21; Page 1345; 2474pp; English.
XX
CC ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
CC invention also encompasses antibodies specific for the secreted proteins,
CC the use of the secreted proteins in drug screening and recombinant
CC vectors and host cells comprising a nucleic acid of the invention. The
CC secreted proteins are thought to be involved in biological activities
CC associated with cellular signalling, cellular differentiation, cell
CC migration, prohormone activation and neurotransmitter activity. The
CC secreted proteins, nucleic acids encoding them, antibodies or antibody
CC fragments specific for the secreted proteins, and modulators of protein
CC activity are useful for diagnosing or treating cancers or other
CC hyperproliferative disorders. Additionally, the secreted proteins and
CC their nucleic acids may also be used in the treatment of autoimmune
CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
CC wound healing. Nucleic acids of the invention may be used for chromosome
CC identification, chromosome mapping, in gene therapy, for identifying
CC individuals from minute biological samples, as hybridisation probes, and
CC as molecular weight markers. The present sequence represents a human
CC secreted protein-encoding cDNA clone of the invention
XX
SQ Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;

Query Match 96.7%; Score 1392.8; DB 7; Length 1460;
Best Local Similarity 99.3%; Pred. No. 7.5e-301;
Matches 1430; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

QY 1 AAAGTAACGGCTACAGACAGTGAAGTAATGTTGCTTCGCGCGGCTGAGAAAACTGTGCG 60
DB 22 AAAGTAACGGCTACAGACAGTGAAGTAATGTTGCTTCGCGCGGCTGAGAAAACTGTGCG 81
QY 61 GTAACCAACCCAGAGCGGTGAGAGCAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 120
DB 82 GTACCAACCCAGAGCGGTGAGAGCAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 141

QY	122	GGACTCAGACTTCA	CCAGCCCACTCGT	TCCAGCCCTTGTA	CGCAAGAGACGCCAAGAC	180	
Db	142	GGACTCAGACTTCA	CCAGCCCACTCGT	TCCAGCCCTTGTA	CGCAAGAGACGCTCAAGAC	201	
QY	181	GCGCTCTCCCGCGT	CCAGGCAACCCCA	GGCTTGCTGGCTT	GGCTGGCTGGCA	240	
Db	202	GCGCTCTCCCGCGT	CCAGGCAACCCCA	GGCTTGCTGGCTT	GGCTGGCTGGCA	261	
QY	241	ACTCGCCCGCGT	GCAGCATGA	CCCTGTGAA	CGCGCTACTGCTTTT	300	
Db	262	ACTCGCCCGCGT	GCAGCATGA	CCCTGTGAA	CGCGCTACTGCTTTT	321	
QY	301	CGGCATG	CCGCAAGCTTCA	GGCTTCCACTGCTCAT	CGTTATTTCTAGTGT	360	
Db	322	CGGCATG	CCGCAAGCTTCA	GGCTTCCACTGCTCAT	CGTTATTTCTAGTGT	381	
QY	361	GCAGCAAGCTT	CCCTGCTCAT	CTTGGCGGGAATCCGT	GGCACTCGCGCTGGTTG	420	
Db	382	GCAGCAAGCTT	CCCTGCTCAT	CTTGGCGGGAATCCGT	GGCACTCGCGCTGGTTG	441	
QY	421	GTGAGAGTTCT	TCTCAGTCTGTTCAT	TAGGCGCAGAAATTG	GGCTGTGCACTTCAGTGCA	480	
Db	442	GTGAGAGTTCT	TCTCAGTCTGTTCAT	TAGGCGCAGAAATTG	GGCTGTGCACTTCAGTGCA	501	
QY	481	GAATGGTTCTG	GGGTACAGTGAACA	CCAAACATCCTCA	AAAGCCTTCAAGCGCAGCGCGC	540	
Db	502	GAATGGTTCTG	GGGTACAGTGAACA	CCAAACATCCTCA	AAAGCCTTCAAGCGCAGCGCGC	561	
QY	541	GTTACAGCCCGT	GTCTGGTCTGTCTG	TCTGTGGCCCTGGA	GGGCATTATATTACACTC	600	
Db	562	GTTACAGCCCGT	GTCTGGTCTGTCTG	TCTGTGGCCCTGGA	GGGCATTATATTACACTC	621	
QY	601	ACCCAGTGCA	TCAAGCTGAACGAGA	CCATTGACTACAACGAG	AGTTCACTGGCGCTG	660	
Db	622	ACCCAGTGCA	TCAAGCTGAACGAGA	CCATTGACTACAACGAG	AGTTCACTGGCGCTG	681	
QY	661	AAAGAGATTAC	GGCCGCGAGTACGGGA	ACGCACTGAGAA	GGGGCTGCCGACCCAGTG	720	
Db	682	AAAGAGATTAC	GGCCGCGAGTACGGGA	ACGCACTGAGAA	GGGGCTGCCGACCCAGTG	741	
QY	721	CTCTACCTGG	CGGAGAAGTTACAC	CCGAGTAGCCCTTG	CGGCTGTACCA	780	
Db	742	CTCTACCTGG	CGGAGAAGTTACAC	CCGAGTAGCCCTTG	CGGCTGTACCA	801	
QY	781	CTGGCGGACA	CTACGCTCGGCCA	CGCTATGGGTGGCTT	GTGCTTCTGCTCTCTCC	840	
Db	802	CTGGCGGACA	CTACGCTCGGCCA	CGCTATGGGTGGCTT	GTGCTTCTGCTCTCTCC	861	
QY	841	AAAGTGTGCT	CTTCCA	CGCCGGCCCGCTCT	ACGGAAGGCTGGCACTG	900	
Db	862	AAAGTGTGCT	CTTCCA	CGCCGGCCCGCTCT	ACGGAAGGCTGGCACTG	921	
QY	901	GCTTTCGCGCT	CTTTCGGGGTCTT	TGACCTTGACCTT	CACTCTAGCGCTTGCCCG	960	
Db	922	GCTTTCGCGCT	CTTTCGGGGTCTT	TGACCTTGACCTT	CACTCTAGCGCTTGCCCG	981	
QY	961	CTCCGCTAG	AGGTCTCCGCGCTA	CCACTCA	GTACGCGCGCTTCT	GGGTACGCTG	1020
Db	982	CTCCGCTAG	AGGTCTCCGCGCTA	CCACTCA	GTACGCGCGCTTCT	GGGTACGCTG	1041
QY	1021	GCAACCGGCGT	CTCTGTGCTCTT	CTCTCGAGGGGCGGT	GTGAGTCTCCAGTATGTTGG	1080	
Db	1042	GCAACCGGCGT	CTCTGTGCTCTT	CTCTCGAGGGGCGGT	GTGAGTCTCCAGTATGTTGG	1101	
QY	1081	CCGAGCGCTCT	TGCAACCTTCTG	GAACCAAAAGCGCA	AGACTGCAAGCCAGAGAGGG	1139	
Db	1102	CCGAGCGCTCT	TGCAACCTTCTG	GAACCAAAAGCGCA	AGACTGCAAGCCAGAGAGGG	1161	
QY	1140	GGGCTCACCTCT	TATTCCTCGGCGA	CCCACTGCA	CAAGCAGGCGCTCTCCAGAC	1198	
Db	1162	GGGCTCACCTCT	TATTCCTCGGCGA	CCCACTGCA	CAAGCAGGCGCTCTCCAGAC	1220	

QY	1199	AATGTATCACCACCTAACCTGTGAGGGGGACCCAAATCTGGA	CTCCTTCCCGCCTTGGGAC	1258
Db	1221	AATGTATCACCACCTAACCTGTGAGGGGGACCCAAATCTGGA	CTCCTTCCCGCCTTGGGAC	1280
QY	1259	ATCCGAGCGCCGGGAAGCAGTGTCCCGCCAGGCGCTGGCGCAGAGA	AGCTCCAGGAAGGGCAC	1318
Db	1281	ATCCGAGCGCCGGGAAGCAGTGTCCCGCCAGGCGCTGGCGCAGAGA	AGCTCCAGGAAGGGCAC	1340
QY	1319	TGAGCGCTGTGGCGCGAGGCGCTCGACAATCCGAGAGCAACAGGAA	AGTCTCTGGGGC	1378
Db	1341	TGAGCGCTGTGGCGCGAGGCGCTCGACAATCCGAGAGCAACAGGAA	AGTCTCTGGGGC	1400
QY	1379	GATCTGTAATAAACCTTTTTTCTTTTGTTTTTAAAAA	AAAAAAAAAAAAAAAAAAAA	1438
Db	1401	GATCTGTAATAAACCTTTTTTCTTTTGTTTTTAAAAA	AAAAAAAAAAAAAAAAAAAA	1460

RESULT	8
ADC20289	
ID	ADC20289 standard; DNA; 1460 BP
XX	
XX	
AC	ADC20289;
XX	
DT	18-DEC-2003 (first entry)

Human secreted protein coding sequence #228.

gene therapy; human; secreted protein; haemopoietic disorder; haematological disorder; anaemia; haemophilia; inflammatory disorder; inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;; leukemia; wound healing; epithelial cell proliferation disorder; immune disorder; autoimmune disorder; asthmatic disorder; cardiovascular disorder; atherosclerosis; myocarditis; infectious disease; HIV; AIDS; endocrine disorder; diabetes; gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.

Homo sapiens.

WO200292787-A2.

21-NOV-2002.

26-MAR-2002; 2002WO-US009257

27-MAR-2001; 2001US-0278650P.

12-SEP-2001; 2001US-00950083.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM;

WPI: 2003-129287/12.

New human secreted proteins and nucleic acid molecules, useful for

preventing or treating hematopoietic disorders, e.g. preparing a diagnostic composition for diagnosing

allenda or nemophilidæ.

Claim 1: SEO ID NO 238; 1512pp; English.

The invention comprises the amino acid and coding sequences of human

useful for detecting, preventing, diagnosing, prognosticating, treating

and haemophilia): inflammatory disorders (e.g., inflammatory bowel diseases

wound healing and disorders of epithelial cell proliferation; immune

cardiovascular disorders (e.g. atherosclerosis and myocarditis);

and gastrointestinal disorders (e.g. duodenal ulcers and infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);

gastroenteritis). The present DNA sequence encodes a human secreted

CC protein of the invention.

XX Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;

Query Match 96.7%; Score 1392.8; DB 9; Length 1460;
Best Local Similarity 99.3%; Pred. No. 7.5e-301;
Matches 1430; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

```

QY 1 AAAGTAACGGCTACAGACAGATGAGAAATAGTTTGGCTGCGCGGCTAGAAAACTCTGTG 60
DB 22 AAAGTAACGGCTACAGACAGATGAGAAATAGTTTGGCTGCGCGGCTAGAAAACTCTGTG 81
QY 61 GTACCAACCCCAAGCGGTGAGAGCAAGCCCACTCCACGCTTCTTAAAGGAGGTGCA 120
DB 82 GTACCAACCCCAAGCGGTGAGAGCAAGCCCACTCCACGCTTCTTAAAGGAGGTGCA 141
QY 121 GGACTCAGACTTCACCAAGCCCACTCGGTCCAGAGCTTGTACGCAAGAGAGAGAGAG 180
DB 142 GGACTCAGACTTCACCAAGCCCACTCGGTCCAGAGCTTGTACGCAAGAGAGAGAGAG 201
QY 181 GCGCTCTCCCGCGTCCAGAGCAAGCCCACTCGGTGCTTGGCTTGGCTGAGAG 240
DB 202 GCGCTCTCCCGCGTCCAGAGCAAGCCCACTCGGTGCTTGGCTTGGCTGAGAG 261
QY 241 ACTCGCGCGCGGTGAGCATGACCTGTGAAAGCGGCTACTGCTTTTAAAGGAGAG 300
DB 262 ACTCGCGCGCGGTGAGCATGACCTGTGAAAGCGGCTACTGCTTTTAAAGGAGAG 321
QY 301 CGGCATGCGCGAGGCTTACGCTTCCACTGCTCATGTTATTTAGTGTGCTCTA 360
DB 322 CGGCATGCGCGAGGCTTACGCTTCCACTGCTCATGTTATTTAGTGTGCTCTA 381
QY 361 GCAGCAAGCTTCTGCTCATGTTGCGGAGATCCGTGAGCACTGCGCTGTTGTTG 420
DB 382 GCAGCAAGCTTCTGCTCATGTTGCGGAGATCCGTGAGCACTGCGCTGTTGTTG 441
QY 421 GTGAGAGTTCTTCTGCTCATGTTGCGGAGATCCGTGAGCACTGCGCTGTTGTTG 480
DB 442 GTGAGAGTTCTTCTGCTCATGTTGCGGAGATCCGTGAGCACTGCGCTGTTGTTG 501
QY 481 GAATGTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
DB 502 GAATGTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 561
QY 541 GTTACAGCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
DB 562 GTTACAGCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 621
QY 601 ACCCCAGTGCATCAGCTGAACGAGACATTGACTTCAACGAGAGTTCACTTGGCGT 660
DB 622 ACCCCAGTGCATCAGCTGAACGAGACATTGACTTCAACGAGAGTTCACTTGGCGT 681
QY 661 AAAGAGATTACCGCGCGGAGTACGCAAGCACTGAGAGAGAGAGAGAGAGAGAGAG 720
DB 682 AAAGAGATTACCGCGCGGAGTACGCAAGCACTGAGAGAGAGAGAGAGAGAGAGAG 741
QY 721 CTCTACCTGGCGGAGAGAGTTCACACCGAGTAGCCCTTGGCGCTGTACCAAGAGT 780
DB 742 CTCTACCTGGCGGAGAGAGTTCACACCGAGTAGCCCTTGGCGCTGTACCAAGAGT 801
QY 781 CTGGCGGAGACTACGCTCGGCCACGCTATAGGTGGCTTCTGCTTCTGCTCTCTCC 840
DB 802 CTGGCGGAGACTACGCTCGGCCACGCTATAGGTGGCTTCTGCTTCTGCTCTCTCC 861
QY 841 AACGTGCTGCTTCCACGCGCGCGCTCTACGAGGCGCTGCACTGCTGACACCGGA 900
DB 862 AACGTGCTGCTTCCACGCGCGCGCTCTACGAGGCGCTGCACTGCTGACACCGGA 921
QY 901 GCCTTGGCGCTCTTGGGAGTCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 960
DB 922 GCCTTGGCGCTCTTGGGAGTCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 981
QY 961 CTCGCGCTAGGCTCTCCGCGCTCACCACTAGTACGCGCGCGCTTCTGCTGCTG 1020

```

```

DB 982 CTCGCGCTAGGCTCTCCGCGCTCACCACTAGTACGCGCGCGCTTCTGCTGCTG 1041
QY 1021 GCAACCGCGCTCTGCTGCTCTCTCTCTCTGAGAGGCGCGTGTGAGTCTCCAGTATGTCG 1080
DB 1042 GCAACCGCGCTCTGCTGCTCTCTCTCTCTGAGAGGCGCGTGTGAGTCTCCAGTATGTCG 1101
QY 1081 CCCAGCGCTCTTCCGACCCCTTCTGAGACCAAGCGCCCAAGAGAGTGC-AGCCAGAGAGAGG 1139
DB 1102 CCCAGCGCTCTTCCGACCCCTTCTGAGACCAAGCGCCCAAGAGAGTGC-AGCCAGAGAGAGG 1161
QY 1140 GGGCTCACTCTTATCTCTGCGGAGCCCACTGCAAGAGAGGCGCTCTCCAGAG-TTAA 1198
DB 1162 GGGCTCACTCTTATCTCTGCGGAGCCCACTGCAAGAGAGGCGCTTCCAGAGCTTTAA 1220
QY 1199 AATGATCAACCACTTACCTGTGAGGAGGAGCCCAATCTGACTCTTCCCGCTTGGAG 1258
DB 1221 AATGATCAACCACTTACCTGTGAGGAGGAGCCCAATCTGACTCTTCCCGCTTGGAG 1280
QY 1259 ATGCGAGGCGCGGAGAGAGTGCAGGCGGAGGCGGCGGAGAGAGTCTCCAGAGAGGAG 1318
DB 1281 ATGCGAGGCGCGGAGAGAGTGCAGGCGGAGGCGGCGGAGAGAGTCTCCAGAGAGGAG 1340
QY 1319 TGAGCGCTGCTGCGGAGAGGCGCTGAGACATCCGAGAGGAGAGAGTCTCTGAGG 1378
DB 1341 TGAGCGCTGCTGCGGAGAGGCGCTGAGACATCCGAGAGGAGAGAGTCTCTGAGG 1400
QY 1379 GATCTGTAATAAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1438
DB 1401 GATCTGTAATAAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1460

```

RESULT 9

AAA94624
ID AAA94624 standard; DNA; 498 BP.

AAA94624;

11-JAN-2001 (first entry)

Human CASB618 EST.

Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian;

colon; autoimmune disease; HLA_A0201; expressed sequence tag; EST; ss.

Homo sapiens.

WO200053748-A2.

14-SEP-2000.

09-MAR-2000; 2000WO-EP002048.

11-MAR-1999; 99GB-00005607.

01-SEP-1999; 99GB-00020590.

(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

Bruck CEM, Cassart J, Coche T, Vinals Y De Bassols;

WPI; 2000-572268/53.

New human CASB618 polypeptide, useful as a vaccine for prophylactic and

therapeutic treatment of cancers, particularly ovarian or colon cancer,

autoimmune diseases and related conditions.

Claim 32; Page 62; 76pp; English.

The present sequence is an expressed sequence tag (EST) for human CASB618

protein. The gene for human CASB618 is thought to be located on

chromosome 15. CASB618 protein and epitopes of CASB618 protein (see

AA026327 to AAB26399) are useful in diagnosing the occurrence of tumour

cells and in vaccines for prophylactic and therapeutic treatment of

CC cancers, particularly ovarian or colon cancer, autoimmune diseases and
CC related conditions

XX Sequence 498 BP; 97 A; 166 C; 137 G; 98 T; 0 U; 0 Other;

Query Match 32.3%; Score 465.4; DB 3; Length 498;
Best Local Similarity 99.4%; Pred. No. 4.4e-94;
Matches 488; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

```
QY 939 CTCTAGGCTGCCGCTCTGCCGCT-CCGCTAGGCTCTCCGGCTCAGCACTAGTACG 997
DB 1 CTCTAGGCTGCCGCTCTGCCGCTCCCGCTAGGCTCTCCGGCTCAGCACTAGTACG 60
QY 998 -GCGCGCCTTCTGGGTACGCTGGGAACCGGCGTCTGTGCTCTTCTCGGAGGGGCC 1056
DB 61 AGCGCGCCTTCTGGGTACGCTGGGAACCGGCGTCTGTGCTCTTCTCGGAGGGGCC 120
QY 1057 GTGTGAGTCTCCAGTATGTTTGGCCGAGCGCTCTTCCGACACCTTCTGACCAAAAGCGCC 1116
DB 121 GTGTGAGTCTCCAGTATGTTTGGCCGAGCGCTCTTCCGACACCTTCTGACCAAAAGCGCC 180
QY 1117 AAGGACTGCAGCCAGAGAGAGGGGCTCACCCTTATCTCTCGGAGACCCCACTGCAAG 1176
DB 181 AAGGACTGCAGCCAGAGAGAGGGGCTCACCCTTATCTCTCGGAGACCCCACTGCAAG 240
QY 1177 CAGCGCCTCTCCCAAGTAAATGATACCACTAACCTGTGAGGGGACCAATCTG 1236
DB 241 CAGCGCCTCTCCCAAGTAAATGATACCACTAACCTGTGAGGGGACCAATCTG 300
QY 1237 GACTCTTCCCGCCTTGGGACATGCAAGCCGGGAGAGAGTCCCGCAGGCTGTGGCC 1296
DB 301 GACTCTTCCCGCCTTGGGACATGCAAGCCGGGAGAGAGTCCCGCAGGCTGTGGCC 360
QY 1297 AGGAGAGCTCCAGAAAGGCACTGAGCGCTGCTGGCGCGAGGCTCGGACATCCGAGGC 1356
DB 361 AGGAGAGCTCCAGAAAGGCACTGAGCGCTGCTGGCGCGAGGCTCGGACATCCGAGGC 420
QY 1357 ACCAGGAAAGTCTCTCTGGGGCGATCTGTAATAAACCTTTTCTTTGTTTTTAA 1416
DB 421 ACCAGGAAAGTCTCTCTGGGGCGATCTGTAATAAACCTTTTCTTTGTTTTTAA 480
QY 1417 AAAAAAAAAA 1427
DB 481 AAAAAAAAAA 491
```

RESULT 10

ABZ68115
ID ABZ68115 standard; DNA; 5033 BP.

XX ABZ68115;

DT 26-MAR-2003 (first entry)

DE Human secreted protein encoding genomic DNA SEQ ID NO 1638.

XX Human; secreted protein; nootropic; neuroprotective; cyostatic;
KM virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KM vulnary; antibacterial; antiparkinsonian; antislaking; antianaemic;
KM antiahrtic; cancer; antirheumatic; hepatotropic; cerebroprotective;
KM antiinflammatory; antiallergic; antidiabetic; antilicer; anticonvulsant;
KM antifungal; antiparasitic; cardian; immune disorder; infection; vaccine;
KM cardiovascular disorder; neurological disease; nephrotropic;
KM gene therapy; gene; ds.

OS Homo sapiens.

PN WO200277186-A2.

XX 03-OCT-2002.

PF 26-MAR-2002; 2002WO-US009188.

XX

PR 27-MAR-2001; 2001US-0278650B.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2003-040583/03.

XX New human secreted proteins encoded by genes contained in cDNA clones
PT (e.g. HGCA19), useful for preventing, treating or diagnosing e.g. AIDS,
PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
PT West Nile fever.

PS Disclosure; Page 2263-2264; 2423pp; English.

XX The invention relates to novel human genes (ABZ6891-ABZ68209) and the
CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections

XX Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;

Query Match 28.1%; Score 405.4; DB 7; Length 5033;
Best Local Similarity 98.6%; Pred. No. 2e-80;
Matches 409; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
QY 1 AAGTAAAGGCTACAGACAGTGAATAATGTTTCGCTCCCGGCTAGAAAAACTGTGCG 60
DB 13 AAGTAAAGGCTACAGACAGTGAATAATGTTTCGCTCCCGGCTAGAAAAACTGTGCG 72
QY 61 GTACCAACCCAGAGCGTTGAGAGAGAGCCCACTCCACGCTTCTTAACGAGAGTGCA 120
DB 73 GTACCAACCCAGAGCGTTGAGAGAGAGCCCACTCCACGCTTCTTAACGAGAGTGCA 132
QY 121 GGAATCAGACTTCACCAAGCCCACTCGGTCCAGCCTTGATACGAAGAAGAGCGCAAGGAC 180
DB 133 GGAATCAGACTTCACCAAGCCCACTCGGTCCAGCCTTGATACGAAGAAGAGCGCAAGGAC 192
QY 181 GCGCTCTCCGCGTCCAGGAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
DB 193 GCGCTCTCCGCGTCCAGGAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 252
QY 241 ACTGGCGCGGCTGAGAGATGACCCCTGTGGAACGGCGTAAGTGCCTTTTACCCCAAGCCC 300
DB 253 ACTGGCGCGGCTGAGAGATGACCCCTGTGGAACGGCGTAAGTGCCTTTTACCCCAAGCCC 312
QY 301 CGGATGCGCGAGGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTT 360
DB 313 CGGATGCGCGAGGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTT 372
QY 361 GGAGCAAGCTTCTGTCTACTTGTGCGGGGATCCGGCCACTCGGCTGCTGCTT 415
DB 373 GGAGCAAGCTTCTGTCTACTTGTGCGGGGATCCGGCCACTCGGCTGCTGCTT 427
```

RESULT 11

ABZ74587
ID ABZ74587 standard; DNA; 5033 BP.

XX ABZ74587;

AC

XX 12-MAY-2003 (first entry)
 DT Secreted protein gene 360 genomic fragment HUFCL31, SEQ ID NO:1734.
 XX
 DE
 XX Human; secreted protein; cancer; tumour; hyperproliferative disorder;
 KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW drug screening; chromosome identification; chromosome mapping;
 KW cytosolic; gene therapy; anti-inflammatory; immunomodulator; anti-HIV;
 KW anti-naemic; vulnary; chromosome 9p21; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200277013-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 26-MAR-2002; 2002WO-US009370.
 XX
 PR 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2003-040578/03.
 XX
 PT New human secreted proteins and nucleic acids, useful for detecting or
 PT treating cancer or other hyperproliferative disorders, autoimmune
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
 XX
 PS Disclosure; Page 2315-2316; 2474pp; English.
 XX
 CC ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
 CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
 CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
 CC invention also encompasses antibodies specific for the secreted proteins,
 CC the use of the secreted proteins in drug screening and recombinant
 CC vectors and host cells comprising a nucleic acid of the invention. The
 CC secreted proteins are thought to be involved in biological activities
 CC associated with cellular signalling, cellular differentiation, cell
 CC migration, prohormone activation and neurotransmitter activity. The
 CC secreted proteins, nucleic acids encoding them, antibodies or antibody
 CC fragments specific for the secreted proteins, and modulators of protein
 CC activity are useful for diagnosing or treating cancers or other
 CC hyperproliferative disorders. Additionally, the secreted proteins and
 CC their nucleic acids may also be used in the treatment of autoimmune
 CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
 CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
 CC wound healing. Nucleic acids of the invention may be used for chromosome
 CC identification, chromosome mapping, in gene therapy, for identifying
 CC individuals from minute biological samples, as hybridisation probes, and
 CC as molecular weight markers. The present sequence represents a human
 CC secreted protein genomic fragment referred to in the disclosure of the
 CC invention
 XX
 SQ Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;
 XX
 QY Query Match 28.1%; Score 405.4; DB 7; Length 5033;
 DB Best Local Similarity 98.6%; Pred. No. 2e-80;
 DB Matches 409; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 1 AAAGTAACGGCTACAGACAGTGAAGAAATAGTTGGCTCGCGGCTAGAAAAAATCTGTGCG 60
 DB 13 AAAGTAACGGCTACAGACAGTGAAGAAATAGTTGGCTCGCGGCTAGAAAAAATCTGTGCG 72
 QY 61 GTACCAACCCCGAGAGCGTTGAGAGAGAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 120
 DB 73 GTACCAACCCCGAGAGCGTTGAGAGAGAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 132

QY 121 GGACTCAGACTTACACGAGCCCACTCGGTCCCAAGCCTTGATACGCAAGAGACGCCAAGAC 180
 DB 133 GGACTCAGACTTACACGAGCCCACTCGGTCCCAAGCCTTGATACGCAAGAGAGCCCAAGAC 192
 QY 181 GCGCTCTCCCGCGCTTCAGAGAGCCCAAGCTTGCTGCTGCCCCGCTGCGTGACAGC 240
 DB 193 GCGCTCTCCCGCGCTTCAGAGAGCCCAAGCTTGCTGCTGCCCCGCTGCGTGACAGC 252
 QY 241 ACTCGGCGCGCGTGAAGCATGACCTGTGGAACGGCGTACTGCTTTTACCCCAAGCCC 300
 DB 253 ACTCGGCGCGCGTGAAGCATGACCTGTGGAACGGCGTACTGCTTTTACCCCAAGCCC 312
 QY 301 CGGCATGCGCGCAGGCTTACGCTTCCACTGCTCATCGTTATCTAGTGTGCTCTA 360
 DB 313 CGGCATGCGCGCAGGCTTACGCTTCCACTGCTCATCGTTATCTAGTGTGCTCTA 372
 QY 361 GCAGCAAGCTTCTGCTCATCTTGCGGGGATCGGTGGCCACTCGCGTGTGTT 415
 DB 373 GCAGCAAGCTTCTGCTCATCTTGCGGGGATCGGTGGCCACTCGGTAAAGGTGT 427

RESULT 12
 ID ADC21005 standard; DNA; 5033 BP.
 AC ADC21005;
 DT 18-DEC-2003 (first entry)
 XX
 DE Human secreted protein-related DNA sequence #423.
 XX
 KW gene therapy; human; secreted protein; haemopoietic disorder;
 KW haematological disorder; anaemia; haemophilia; inflammatory disorder;
 KW inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
 KW leukaemia; wound healing; epithelial cell proliferation disorder;
 KW immune disorder; autoimmune disorder; asthmatic disorder;
 KW cardiovascular disorder; atherosclerosis; myocarditis;
 KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;
 KW gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200292787-A2.
 PD 21-NOV-2002.
 PF 26-MAR-2002; 2002WO-US009257.
 PR 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2003-129287/12.
 XX
 PT New human secreted proteins and nucleic acid molecules, useful for
 PT preparing a diagnostic or pharmaceutical composition for diagnosing,
 PT preventing or treating hematopoietic or hematologic disorders, e.g.
 PT anemia or hemophilia.
 XX
 PS Disclosure; SEQ ID NO 959; 1512pp; English.
 XX
 CC The invention comprises the amino acid and coding sequences of human
 CC secreted proteins. The DNA and protein sequences of the invention are
 CC useful for detecting, preventing, diagnosing, prognosticating, treating
 CC or ameliorating: haematopoietic or haematological disorders (e.g. anaemia
 CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease
 CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
 CC wound healing and disorders of epithelial cell proliferation; immune
 CC disorders (e.g. autoimmune disorders and asthmatic disorders);

CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);
CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
CC and gastrointestinal disorders (e.g. duodenal ulcers and
CC gastroenteritis). The present DNA sequence was used in the
CC exemplification of the invention.

XX Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;

Query Match 28.1%; Score 405.4; DB 9; Length 5033;
Best Local Similarity 98.6%; Pred. No. 2e-80;
Matches 409; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAAGTAACGGCTACAGACAGTGAAGAAATAGTTTGGCTCGCCGCTAGAAAACTCTGTG 60
DB 13 AAAGTAACGGCTACAGACAGTGAAGAAATAGTTTGGCTCGCCGCTAGAAAACTCTGTG 72
QY 61 GTACCAACCCAGAGCGTTGAGAGCAAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 120
DB 73 GTACCAACCCAGAGCGTTGAGAGCAAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 132
QY 121 GGACTCAGACTTCAACCAAGCCCACTCGGTCCAGCCTTGTACGCAAGAGACGCCAAGAC 180
DB 133 GGACTCAGACTTCAACCAAGCCCACTCGGTCCAGCCTTGTACGCAAGAGACGCCAAGAC 192
QY 181 GCGCTCTCCCGCGTCCAGAGCAGCCCAAGCTTGTGCTTGCCTGCCCGCTGGTGACG 240
DB 193 GCGCTCTCCCGCGTCCAGAGCAGCCCAAGCTTGTGCTTGCCTGCCCGCTGGTGACG 252
QY 241 ACTCGCGCGCGGTGACAGATGACCCCTGTGAACGGCGTACTGCTTTTACCCCAAGCC 300
DB 253 ACTCGCGCGCGGTGACAGATGACCCCTGTGAACGGCGTACTGCTTTTACCCCAAGCC 312
QY 301 CGGCATGCCGAGCGTTGAGCGTTCCACTGCTCATCGTTATTCTAGTGTGTTGGCTCTA 360
DB 313 CGGCATGCCGAGCGTTGAGCGTTCCACTGCTCATCGTTATTCTAGTGTGTTGGCTCTA 372
QY 361 GCAGCAAGCTTCTGCTCATCTTTGCGCGGAGTCCGTGGCCACTCGCGCTGTTT 415
DB 373 GCAGCAAGCTTCTGCTCATCTTTGCGCGGAGTCCGTGGCCACTCGGTAAAGGTGT 427

RESULT 13
ADD19187

ID ADD19187 standard; cDNA; 580 BP.

XX ADD19187;

DT 15-JAN-2004 (first entry)

DE Human cDNA from secreted protein gene 4.

XX human secreted protein; cytostatic; antibacterial; virucide;
KW neuroprotective; gynaecological; gastrointestinal-Gen; cardiant;
KW cardiovascular-Gen; nephrotropic; antiinflammatory; muscular-Gen;
KW respiratory-Gen; immunosuppressive; cerebroprotective; vasotropic;
KW neotropic; antiallergic; cancer; bacterial infection; viral infection;
KW neural disorder; immune system disorder; blood disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW inflammatory disorder; proliferative disorder; human; ss; gene.

OS Homo sapiens.

PN WO2003052377-A2.

PD 26-JUN-2003.

PF 06-NOV-2002; 2002WO-US035606.

PR 07-NOV-2001; 2001US-0331046P.

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Ruben SM;
XX WPI; 2003-533050/50.
DR P-PSDB; ADD19262.

XX New isolated nucleic acids encoding signal transduction pathway component
PT polypeptides, useful for diagnosing, treating, and/or preventing
PT disorders, such as cancer, infections, cardiovascular and inflammatory
PS diseases.

PS Claim 1; SEQ ID NO 14; 554bp; English.

XX The invention relates to an isolated nucleic acid molecule (cDNA)
CC encoding a human secreted protein, representing one of 85 novel genes.
CC Also included are recombinant vectors, host cells (expressing the
CC protein), the secreted proteins (including their fragments, epitopes and
CC homologues), an isolated antibody that binds specifically to the protein,
CC diagnosing a pathological condition or susceptibility to a pathological
CC condition (comprising determining the presence or absence of a mutation
CC in the nucleic acid and diagnosing a pathological condition or
CC absence of the mutation), diagnosing a pathological condition or
CC susceptibility to a pathological condition (comprising determining the
CC presence or amount of expression of the protein in a biological sample
CC and diagnosing a condition based on the presence or amount of expression
CC of the protein), preventing, treating or ameliorating a medical condition
CC by administering the nucleic acid or protein to a mammalian subject,
CC identifying a binding partner to the protein, the gene corresponding to
CC the cDNA sequence, and identifying an activity in a biological assay
CC (comprising expressing the nucleic acid in a cell, isolating the
CC supernatant, detecting an activity in a biological assay and identifying
CC the protein in the supernatant having the activity). The nucleic acids
CC and proteins display the following activities: cytostatic, antibacterial,
CC virucide, neuroprotective, gynaecological, gastrointestinal-Gen,
CC Cardiant, Cardiovascular-Gen, Nephrotropic, Antiinflammatory, Muscular-
CC Gen, Respiratory-Gen, Immunosuppressive, Cerebroprotective, Vasotropic,
CC Nootropic, Antiallergic. The methods and compositions of the present
CC invention are useful for diagnosing, treating, preventing and/or
CC prognosticating disorders related to the novel polypeptides, such as
CC cancer, bacterial or viral infections, and neural, immune system, blood,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal, inflammatory or proliferative disorders (many examples of these
CC diseases and disorders are given in the specification). The present
CC sequence encodes a novel secreted protein of the invention.

XX Sequence 580 BP; 115 A; 198 C; 156 G; 111 T; 0 U; 0 Other;

Query Match 28.1%; Score 405; DB 9; Length 580;
Best Local Similarity 100.0%; Pred. No. 1.4e-80;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTAACGGCTACAGACAGTGAAGAAATAGTTTGGCTCGCCGCTAGAAAACTCTGTG 60
DB 13 AAAGTAACGGCTACAGACAGTGAAGAAATAGTTTGGCTCGCCGCTAGAAAACTCTGTG 72
QY 61 GTACCAACCCAGAGCGTTGAGAGCAAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 120
DB 73 GTACCAACCCAGAGCGTTGAGAGCAAGCCCACTCCACGCTTCTTAACGAGAGGTGCA 132
QY 121 GGACTCAGACTTCAACCAAGCCCACTCGGTCCAGCCTTGTACGCAAGAGACGCCAAGAC 180
DB 133 GGACTCAGACTTCAACCAAGCCCACTCGGTCCAGCCTTGTACGCAAGAGACGCCAAGAC 192
QY 181 GCGCTCTCCCGCGTCCAGAGCAGCCCAAGCTTGTGCTTGCCTGCCCGCTGGTGACG 240
DB 193 GCGCTCTCCCGCGTCCAGAGCAGCCCAAGCTTGTGCTTGCCTGCCCGCTGGTGACG 252
QY 241 ACTCGCGCGCGGTGACAGATGACCCCTGTGAACGGCGTACTGCTTTTACCCCAAGCC 300
DB 253 ACTCGCGCGCGGTGACAGATGACCCCTGTGAACGGCGTACTGCTTTTACCCCAAGCC 312
QY 301 CGGCATGCCGAGCGTTGAGCGTTCCACTGCTCATCGTTATTCTAGTGTGTTGGCTCTA 360
DB 313 CGGCATGCCGAGCGTTGAGCGTTCCACTGCTCATCGTTATTCTAGTGTGTTGGCTCTA 372

CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 2684 BP, 599 A; 785 C; 684 G; 616 T; 0 U; 0 Other;

Query Match 27.5%; Score 396; DB 7; Length 2684;
Best Local Similarity 97.6%; Pred. No. 2.1e-78;
Matches 402; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
QY 1007 TCTGGGTACCGCTGGCAACCGCGCTGTGCTTCTTCTCGGAGGGCCGTGGTGAATC 1066
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2363 TTTCGATCCCCACCGCCACAGCGCTGCTGCTTCTTCTCGGAGGGCCGTGGTGAATC 2304

QY 1067 TCCAGTATGTTGGCCCGCGCTCTTGGCAACCTTCTGGACCAAGCGCCAGAGACTGCA 1126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2303 TCCAGTATGTTGGCCCGCGCTCTTGGCAACCTTCTGGACCAAGCGCCAGAGACTGCA 2244

QY 1127 GCCAGAGAGAGGGGGCTCACTCTTATCCTCGGCGAACCACTGACACAAGCAGCGCGCTC 1186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2243 GCCAGAGAGAGGGGGCTCACTCTTATCCTCGGCGAACCACTGACACAAGCAGCGCGCTC 2184

QY 1187 TCCCAGACTTAAATGTATCACCACCTAACCCTGTGAGGGGGACCAATCTGACTCCTTCC 1246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2183 TCCCAGACTTAAATGTATCACCACCTAACCCTGTGAGGGGGACCAATCTGACTCCTTCC 2124

QY 1247 CCGCCTTGGGACATCGCAGCGCCGGAAGCAGTGCCCGCCAGGCTGGGCCAGAGAGCTC 1306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2123 CCGCCTTGGGACATCGCAGCGCCGGAAGCAGTGCCCGCCAGGCTGGGCCAGAGAGCTC 2064

QY 1307 CAGGAAGGGCACTGAGCGCTGTGGCGAGAGGCTCGGACATCCGACAGCACAGGGA 1366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2063 CAGGAAGGGCACTGAGCGCTGTGGCGAGAGGCTCGGACATCCGACAGCACAGGGA 2004

QY 1367 GTCTCCTGGGGCGATCTGTAATAAACCTTTTCTTTGTTTAA 1418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2003 GTCTCCTGGGGCGATCTGTAATAAACCTTTTCTTTGTTTAA 1952
```

Search completed: February 23, 2004, 19:50:04
Job time : 586.643 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 19:35:58 ; Search time 130.797 Seconds
(without alignments)
6113.919 Million cell updates/sec

Title: US-09-936-456-1

Perfect score: 1441
Sequence: 1 aaagtaacgctacagacag.....aaaaaaaaaaaaaaaa 1441

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	4.2	7218	1	US-08-232-463-14 Sequence 14, Appl
2	50.2	3.5	1098	4	US-09-252-991A-1304 Sequence 1304, Ap
3	50.2	3.5	1257	4	US-09-252-991A-1255 Sequence 1255, Ap
4	50.2	3.5	1746	4	US-09-252-991A-1163 Sequence 1163, Ap
5	48.2	3.3	396	4	US-09-252-991A-1350 Sequence 1350, Ap
6	48	3.3	1436	4	US-09-614-912-35 Sequence 35, Appl
7	47.6	3.3	420	4	US-09-252-991A-11029 Sequence 11029, A
8	47.6	3.3	1683	4	US-09-252-991A-11226 Sequence 11226, A
9	47.6	3.3	1974	4	US-09-252-991A-11111 Sequence 11111, A
10	47.6	3.3	2283	4	US-09-252-991A-10956 Sequence 10956, A
11	46.4	3.2	1431	4	US-09-614-912-49 Sequence 49, Appl
12	45	3.1	1926	4	US-09-249-585A-4 Sequence 4, Appl
13	45	3.1	1931	2	US-09-130-114-2 Sequence 2, Appl
14	44.4	3.1	1835	4	US-09-485-549-1 Sequence 1, Appl
15	44.4	3.1	152331	3	US-09-128-155-16 Sequence 16, Appl
16	43.8	3.0	462	4	US-09-252-991A-1610 Sequence 1610, Ap
17	43.8	3.0	1362	4	US-09-252-991A-1556 Sequence 1556, Ap
18	43.8	3.0	1551	4	US-09-252-991A-12931 Sequence 12931, A
19	43.8	3.0	1587	4	US-09-252-991A-1561 Sequence 1561, Ap
20	43.8	3.0	1659	4	US-09-252-991A-13081 Sequence 13081, A
21	43.8	3.0	1917	4	US-09-252-991A-1465 Sequence 1465, Ap
22	43.8	3.0	2073	4	US-09-252-991A-12858 Sequence 12858, A
23	43.8	3.0	5852	1	US-07-867-106-2 Sequence 2, Appl
24	43.6	3.0	3722	4	US-10-164-595-9 Sequence 9, Appl
25	43.6	3.0	3862	4	US-10-164-595-5 Sequence 5, Appl
26	43.6	3.0	3937	4	US-10-164-595-7 Sequence 7, Appl
27	43.6	3.0	3985	4	US-10-164-595-3 Sequence 3, Appl

28	43.4	3.0	597	4	US-09-252-991A-5066 Sequence 5066, Ap
29	43.4	3.0	1308	4	US-09-252-991A-5127 Sequence 5127, Ap
30	43.4	3.0	1782	4	US-09-252-991A-5094 Sequence 5094, Ap
31	43.2	3.0	561	4	US-09-252-991A-13572 Sequence 13572, A
32	43.2	3.0	4131	4	US-09-252-991A-13773 Sequence 13773, A
33	43.2	3.0	8211	4	US-09-252-991A-13656 Sequence 13656, A
34	43	3.0	835	6	5223425-7 Patent No. 5223425
35	43	3.0	1093	6	5223425-3 Patent No. 5223425
36	43	3.0	2089	4	US-09-291-922-25 Sequence 25, Appl
37	43	3.0	2196	4	US-09-673-395A-548 Sequence 548, Ap
38	42.8	3.0	1454	4	US-09-372-422A-19 Sequence 19, Appl
39	42.6	3.0	367	3	US-09-328-111-446 Sequence 446, Ap
40	42.6	3.0	744	4	US-09-252-991A-5564 Sequence 5564, Ap
41	42.6	3.0	1332	4	US-09-252-991A-5601 Sequence 5601, Ap
42	42.6	3.0	1681	4	US-09-434-288-7 Sequence 7, Appl
43	42.4	2.9	68750	3	US-09-335-409-1 Sequence 1, Appl
44	42.4	2.9	68750	4	US-09-568-102-1 Sequence 1, Appl
45	42.4	2.9	68750	4	US-09-567-969-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZgpt-F1s
US-08-232-463-14
Query Match 4.2%; Score 61; DB 1; Length 7218;

Best Local Similarity 1.9%; Pred. No. 5.5e-05;
Matches 7; Conservative 221; Mismatches 131; Indels 0; Gaps 0;

```
QY 753 CCCTGGGCTGTACCAACAGTACCACTGCGGAGACACTAGCCTCGCCAGCTATG 812
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1084 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1143
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 813 GGTGGCTTCTGCTTGTGCTCTCTCCACAGTGTCTCTCCAGCCGCGCCCTCTA 872
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1144 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1203
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 873 CGAGGCTTGACACTGTACCAACCGAGCCTTGGCTCTTGGGGTCTTGGCCTTGC 932
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1204 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1263
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 933 CTCATCTCTAGCGTGGCTCTGCGGCTCGCCTAGGCTCTCTCGGCTGACCACTCA 992
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1264 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1323
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 993 GTACGGCGCGCCTTCTGGGTACAGCTGGCAACCGGCTCTGTGCTTCTCTCGAGG 1052
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1324 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1383
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1053 GGCCGTGTGAGTCTCCAGTATGTGGCCCGAGCGCTCTTGCACCTTCTGACCAAA 1111
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1384 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1442
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 2

US-09-252-991A-1304/C
; Sequence 1304, Application US/09252991A

```
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1304
; LENGTH: 1098
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1304
```

Query Match 3.5%; Score 50.2; DB 4; Length 1098;
Best Local Similarity 50.6%; Pred. No. 0.011;
Matches 121; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

```
QY 830 GGCTCTCTCCAAAGTGTCTCTCCAGCCCGCCGCTCTACGAGGCTTGCACTGC 889
      |||||
Db 602 GACTCTCTCAATGCGCTGCGCGCGGTGCGGAGCTGTGCGGCGCGCTGATGTGC 543
      |||||
QY 890 TGACCAACCGAGCCTTCCGCTCTTGGGGTCTTGGCTTGGCTTCATCTAGCGTGC 949
      |||||
Db 542 TGGCGCGCGGCTCGGCGCAACGCGGAGCCTCGCCCTGCGCTGACACACCGCGGCG 483
      |||||
QY 950 CGCTCTGCGCGCTCGCCTAGGCTCTCTCGGCTCACCACTAGTACGGCGCGCTTCT 1009
      |||||
Db 482 TGCTCGCGCGCTGTTCGCGAAGCCTGAGAACACCCCGCGAGCCCGCGGAGGCT 423
      |||||
QY 1010 GGGTACGCTGCAACCGGCGTCTGTGCTCTTCTCTGGAGGCGCGTGTGAGTCTC 1068
      |||||
Db 422 TGCGCCTGCGCGAGCGGAGCCTGCGGCAATTGCTATGACACCTGCGGCGGCTC 364
      |||||
```

RESULT 3
US-09-252-991A-1255

; Sequence 1255, Application US/09252991A
; Patent No. 6551795

```
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONA
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1255
; LENGTH: 1257
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1255
```

Query Match 3.5%; Score 50.2; DB 4; Length 1257;
Best Local Similarity 50.6%; Pred. No. 0.011;
Matches 121; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

```
QY 830 GGCTCTCTCCAAAGTGTCTCTCCAGCCCGCCGCTCTACGAGGCTTGCACTGC 889
      |||||
Db 333 GACTCTCTCAATGCGCTGCGCGGTGCGGAGCTGTGCGGCGCGCTGATGTGC 392
      |||||
QY 890 TGACCAACCGAGCCTTCCGCTCTTGGGGTCTTGGCTTGGCTTCATCTAGCGTGC 949
      |||||
Db 393 TGGCGCGCGGCTCGGCGCAACGCGGAGCCTCGCCTGCGCTGACACACCGCGGCG 452
      |||||
QY 950 CGCTCTGCGCGCTCGCCTAGGCTCTCTCGGCTCACCACTAGTACGGCGCGCTTCT 1009
      |||||
Db 453 TGCTCGCGCGCTGTTCGCGAAGCCTGAGAACACCCCGCGAGCCCGCGGAGGCT 512
      |||||
QY 1010 GGGTACGCTGCAACCGGCGTCTGTGCTCTTCTCTGGAGGCGCGTGTGAGTCTC 1068
      |||||
Db 513 TGCGCCTGCGCGAGCGGAGCCTCGCGCAATTGCTATGACACCTGCGGCGGCTC 571
      |||||
```

RESULT 4

US-09-252-991A-1163
; Sequence 1163, Application US/09252991A

```
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONA
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1163
; LENGTH: 1746
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1163
```

Query Match 3.5%; Score 50.2; DB 4; Length 1746;
Best Local Similarity 50.6%; Pred. No. 0.013;
Matches 121; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

```
QY 830 GGCTCTCTCCAAAGTGTCTCTCCAGCCCGCCGCTCTACGAGGCTTGCACTGC 889
      |||||
Db 1283 GACTCTCTCAATGCGCTGCGCGGTGCGGAGCTGTGCGGCGCGCTGATGTGC 1342
      |||||
QY 890 TGACCAACCGAGCCTTCCGCTCTTGGGGTCTTGGCTTGGCTTCATCTAGCGTGC 949
      |||||
```

Db 1343 TGGCGCGCGGCTCGGGCCGAACGCCGGACCTCGCCCTGCGCTGCACACCGCGCGG 1402
QY 950 CGCTCTGCGCGCTCCGCTAGGCTCTCCGCGCTACACCTACAGTAACGGCGCGCTTCT 1009
Db 1403 TGCTCGCGCGCTGTTCGCGGAAGCCCTGGAGAACACCCCGCGCGCGCGCGAGGCT 1462
QY 1010 GGGTCACGCTGACACCGCGCTCTGCTCTCTCTCGAGGGCGCTGTAGTCTC 1068
Db 1463 TGGCGCTGCGCGCGAGCGGACGCTCGCGGCTTGCCTATGACACCTCTGCGGGGCTC 1521

RESULT 5

US-09-252-991A-1350/c
Sequence 1350, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 1350
LENGTH: 396
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1350

Query Match 3.3%; Score 48.2; DB 4; Length 396;
Best Local Similarity 52.2%; Pred. No. 0.024;
Matches 107; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 830 GGCTCTCTCCACGCTGCTCTCTCCAGCGCGCGCGCTCTACGAGGCGCTGGCACTGC 889
Db 217 GACTCTGCTCAATGCGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 158
QY 890 TGACCAACGAGGAGCTTGGCGCTCTTGGAGGCTTGGCGCTTGGCGCTTGGCGCTTGGCG 949
Db 157 TGGCGCGCGCGCTCGGGCCGAACGCCGGAGCCCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 98
QY 950 CGCTCTGCGCGCTCCGCTAGGCTCTCTCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 1009
Db 97 TGCTCGCGCGCTGTTCGCGGAAGCCCTGGAGAACACCCCGCGCGAGCCCGCGAGGCT 38
QY 1010 GGGTCACGCTGACACCGCGCTCT 1034
Db 37 TGGCGCTGCGCGCGAGCGGAGCGCT 13

RESULT 6

US-09-614-912-35
Sequence 35, Application US/09614912
Patent No. 6677502
GENERAL INFORMATION:
APPLICANT: Allen, Steve
APPLICANT: Rafalski, Antoni
APPLICANT: Orozco, Buddy
APPLICANT: Miao, Gou-Hau
APPLICANT: Famodu, Omolayo O.
APPLICANT: Lee, Jian Ming
APPLICANT: Sakai, Hajime
APPLICANT: Weng, Zude
APPLICANT: Cai, Perry G
APPLICANT: Anderson, Shawn
TITLE OF INVENTION: Plant Metabolism Genes
FILE REFERENCE: B1378 US NA
CURRENT APPLICATION NUMBER: US/09/614,912
CURRENT FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: 60/143,401
PRIOR FILING DATE: 1999-07-12
PRIOR APPLICATION NUMBER: 60/143,412
PRIOR FILING DATE: 1999-07-12
PRIOR APPLICATION NUMBER: 60/146,650
PRIOR FILING DATE: 1999-07-30
PRIOR APPLICATION NUMBER: 60/170,906
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: 60/172,959
PRIOR FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: 60/172,946
PRIOR FILING DATE: 1999-12-21
NUMBER OF SEQ ID NOS: 204
SOFTWARE: Microsoft Office 97
SEQ ID NO 35
LENGTH: 1436
TYPE: DNA
ORGANISM: Zea mays
US-09-614-912-35

Query Match 3.3%; Score 48; DB 4; Length 1436;
Best Local Similarity 46.8%; Pred. No. 0.039;
Matches 187; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

QY 541 GTTACAGCCCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db 43 GCTTCT 102
QY 601 ACCCCAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 103 TTGACAGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 162
QY 661 AAGAGATTAACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Db 163 GACGAGAGCCCAACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 222
QY 721 CTCTACCTGCGGAGAGTTCACACCG--AGTAGCCCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 777
Db 223 GGGGAGCGCGCGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 282
QY 778 CACTGCGGAGCACTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 837
Db 283 CAGGTGTGGGAGCAAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 342
QY 838 TCCACGTGCTGCTCTCCAGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 897
Db 343 GCCTTCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 402
QY 898 GAGCCTTGGCGCTCTTGGGCTCTTGGGCTCTTGGGCTCTTGGGCTCTTGGGCTCTTGGGCTCTTGG 937
Db 403 TGGGCTTACGCGAGAGCTTTCACGCGCGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCT 442

RESULT 7

US-09-252-991A-11029/c
Sequence 11029, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 11029
LENGTH: 420
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-11029

Query Match 3.3%; Score 47.6; DB 4; Length 420;
Best Local Similarity 48.2%; Pred. No. 0.033;
Matches 134; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 785 CCGGACACTACGCGCTCGGCCACGCTATGGGTGGCGTTCTGCTTCTGGCTCTCTCCACAG 844
DB 382 CCGTGGAACACGCTTCTCTCATGCGCTGGCTGGCACTGCACATGATCCACTCGCTGG 323
QY 845 TGCTGCTCTCCACGCGCGCCCGCTCTACGAGGCTTGCACTGCTGACCAACCGAGCT 904
DB 322 CGGTGACCGAGAACGCGCGCGGTGTCAAGAGCTGAGACCGGTCTTCTGGCGATCGCGCTT 263
QY 905 TCGGCTCTTCGGGGTCTTGGCTTGGCTTCACTCTAGCGTGGCGCTGCGCTCC 964
DB 262 TCTCCCTGAGCGCTGTGGGACCTTCTGTGCTCGCTCCGGCGTGTCACTCGGTGACG 203
QY 965 GCCTAGGCTCTCCGCGCTCAGCACTCAGTACGCGCGCGCTTCTGGTCAAGCTGGCAA 1024
DB 202 CTTGCGCTCCGACCGCGAGCGCGGGGTGTTCATCTCGCTTCTGCTGCTGTGTGTCG 143
QY 1025 CCGGCGTCTGTGCTCTTCTCTCGAGAGGGCGCTGTG 1062
DB 142 GCGGCTCGCTGACCTGTGTGCGCTTGTGCGCGCGCGGTG 105

RESULT 8

US-09-252-991A-11226
; Sequence 11226, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11226
; LENGTH: 1683
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11226

Query Match 3.3%; Score 47.6; DB 4; Length 1683;
Best Local Similarity 48.2%; Pred. No. 0.051;
Matches 134; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 785 CCGGACACTACGCGCTCGGCCACGCTATGGGTGGCGTTCTGCTTCTGGCTCTCTCCACAG 844
DB 864 CCGTGGAACACGCTTCTCTCATGCGCTGGCTGGCACTGCACATGATCCACTCGCTGG 923
QY 845 TGCTGCTCTCCACGCGCGCCCGCTCTACGAGGCTTGCACTGCTGACCAACCGAGCT 904
DB 924 CCGTGACCGAGAACGCGCGCGGTGTCAAGAGCTGAGACCGGTCTTCTGGCGATCGCGCTT 983
QY 905 TCGGCTCTTCGGGGTCTTGGCTTGGCTTCACTCTAGCGTGGCGCTGCGCTCC 964
DB 984 TCTCCCTGAGCGCTGTGGGACCTTCTGTGCTCGCTCCGGCGTGTCACTCGGTGACG 1043
QY 965 GCCTAGGCTCTCCGCGCTCAGCACTCAGTACGCGCGCGCTTCTGGTCAAGCTGGCAA 1024
DB 1044 CTTGCGCTCCGACCGCGAGCGCGGGGTGTTCATCTCGCTTCTGCTGTGTGTGTCG 1103
QY 1025 CCGGCGTCTGTGCTCTTCTCTCGAGAGGGCGCTGTG 1062
DB 1104 GCGGCTCGCTGACCTGTGTGCGCTTGTGCGCGCGCGGTG 1141

RESULT 9

US-09-252-991A-11111
; Sequence 11111, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11111
; LENGTH: 1974
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11111

Query Match 3.3%; Score 47.6; DB 4; Length 1974;
Best Local Similarity 48.2%; Pred. No. 0.054;
Matches 134; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 785 CCGGACACTACGCGCTCGGCCACGCTATGGGTGGCGTTCTGCTTCTGGCTCTCTCCACAG 844
DB 731 CCGTGGAACACGCTTCTCTCATGCGCTGGCTGGCACTGCACATGATCCACTCGCTGG 790
QY 845 TGCTGCTCTCCACGCGCGCCCGCTCTACGAGGCTTGCACTGCTGACCAACCGAGCT 904
DB 791 CCGTGACCGAGAACGCGCGCGGTGTTCAGAGCTGAGACCGGTCTTCTGGCGATCGCGCTT 850
QY 905 TCGGCTCTTCGGGGTCTTGGCTTGGCTTCACTCTAGCGTGGCGCTTCTGCGCTCC 964
DB 851 TCTCCCTGAGCGCTGTGGGACCTTCTGTGCTCGCTCCGGCGTGTCACTCGGTGACG 910
QY 965 GCCTAGGCTCTCCGCGCTCAGCACTCAGTACGCGCGCGCTTCTGGTCAAGCTGGCAA 1024
DB 911 CTTGCGCTCCGACCGCGAGCGCGGGGTGTTCATCTCGCTTCTGCTGTGTGTGTCG 970
QY 1025 CCGGCGTCTGTGCTCTTCTCTCGAGAGGGCGCTGTG 1062
DB 971 GCGGCTCGCTGACCTGTGTGCGCTTGTGCGCGCGCGGTG 1008

RESULT 10

US-09-252-991A-10956/c
; Sequence 10956, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10956
; LENGTH: 2283
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10956

Query Match 3.3%; Score 47.6; DB 4; Length 2283;
Best Local Similarity 48.2%; Pred. No. 0.056;
Matches 134; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY	1065	TCCTCAGTATGTTGGGCCCCAGCGCTCTTCGCACCCCTTTGGACCAAGGCCAAGGACTG	1124
Db	946	CCCCGTCTCTCTCTCCCCGTCTCTGTCTCTCAACTCCGGCCCCAGCTCTTCCTCACCCT	1005
QY	1125	CAGCCAGAGAGAGGGGGCTCACCTTTATCTCTGGCGACCCCACTGCACAAGCAGGCCGC	1184
Db	1006	CCGGCCCCAGCTCCTCCATCACCTCCGGCCCCAGCTCTCCATCACCTCCGGGGGCCCA	1065
QY	1185	TCCTCCAGACTTAAATGATATCAACAATAACTGTGAGGGGAGCCCAATCTGACTCCTT	1244
Db	1066	TCCTCTGCACCTTCTCGGTCCGCCCTTCAAGCACTTCTCGGTCCGCCCTCTCCAGCACCT	1125
QY	1245	CCCCGCCCTT	1253
Db	1126	GCACCTCTT	1134

```

RESULT 13
US-09-130-114-2
; Sequence 2, Application US/09130114
; Patent No. 5976807
;
GENERAL INFORMATION:
;
APPLICANT: Horlick, Robert A.
APPLICANT: Damaj, Bassam B.
APPLICANT: Robbins, Alan K.
TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
TITLE OF INVENTION: From Multiple Transfected Episomes
FILE REFERENCE: 0867/1D903US1
CURRENT APPLICATION NUMBER: US/09/130,114
CURRENT FILING DATE: 1998-08-06
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
;
LENGTH: 1931
;
TYPE: DNA
;
ORGANISM: EBNA
US-09-130-114-2

```

Query Match	3.1%;	Score 45;	DB 2;	Length 1931;
Best Local Similarity	44.1%;	Pred. No. 0.22;		
Matches 189;	Conservative 0;	Mismatches 240;	Indels 0;	Gaps 0;

QY	825	CTTCTGGGCTCCTTCCAAAGTGTGCTCTCCACGCGCCGCCCGCTCTACGAGGAGCTGGC	884
Db	706	CTGTGTCCTCCCGGTCTGTCTCTCCCGTCTGTGTCTCTCCCGTCTCTCCCGTCTCGT	765
QY	885	ACTGTGACCAACCGGAGGCTTTCGGGCTCTTGGGGTCTTTCGCTTGGCCATCTCTAG	944
Db	766	CTTCCCGTCTTCCCGTCTGTGTCTCTCCCGTCTCCCGTCTCTGTCTCTCTCCCGT	825
QY	945	CGTGCCTCTTGCCTGCTCCGCTTAGGCTCTTCCGCTACCACTCAGTACGCGCCGC	1004
Db	826	CTTCCCGTCTGTGTCTCTCTCCCGTCTCTCCCGTCTGTGTCTCTCTCCCGTCTCCCGT	885
QY	1005	CTTCTGGGTCACGCTGGCAACCGGGGCTCTGTGTCTCTTCTTGGAGGGGCGGTGTAG	1064
Db	886	CTGTGTCTCTCCCGTCTCTCCCGTCTGTGTCTCTCTCTCCCGTCTCTGTGTCTCTCT	945
QY	1065	TCTTCAGTATGTTGGGCCCCAGCGCTCTTGGCACCCTTCTGGACCAAGGCCAAGGACTG	1124
Db	946	CCCGTCTCTGTCTCTCCCGTCTGTGTCTCTCACACTCCGGCCCAAGCTCTCTGTCACT	1005
QY	1125	CAGCCAGGAGAGAGGGGGCTCACTCTTATCTCTCGGCGAACCCACTGCACAGAGGCGCG	1184
Db	1006	CCGGCCCCAGCTCTTCCATCACTCTCGGGCCCAAGCTCTTCCATCACTCTCGGGGGCCCCA	1065
QY	1185	TCTCCAGACTTAAATATGTATCACCACTAACCTGTGAGGGGGAGCCCAATCTGCACTCCTT	1244
Db	1066	TCTCTGCACTTTCGTGGTCCCCCTTCAAGCACTTTCGTGGTCCCCCTCTTCAAGCACT	1125
QY	1245	CCCGGCTT 1253	

Db 1126 GCACCTCTT 1134

```

RESULT 14
US-09-485-549-1
; Sequence 1, Application US/09485549
; Patent No. 6361948
; GENERAL INFORMATION:
; APPLICANT: James Tricoli
; APPLICANT: Rachel Rhondinelli
; APPLICANT: Fox Chase Cancer Center
; TITLE OF INVENTION: Prognostic Compositions for Prostate Cancer and Methods of U
; TITLE OF INVENTION: thereof
; FILE REFERENCE: PCCC 96-13
; CURRENT APPLICATION NUMBER: US/09/485,549
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/16768
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/055,285
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1835
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-485-549-1

```

Query Match	3.1%;	Score 44.4;	DB 4;	Length 1835;
Best Local Similarity	59.5%;	Pred. No. 0.3;		
Matches	75;	Conservative	0;	Mismatches 51;
			Indels	0;
			Gaps	0;

[illegible]

```

RESULT 15
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; EARLIER FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

```

Query Match 3.18; Score 44.4; DB 3; Length 152331;

Best Local Similarity 51.5%; Pred. No. 1.2;
Matches 102; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
QY 832 CTCCTCTCCAGCTGCTGCTCTCCAGCGCGGCCCGCTCTACGAGGCTTGCACTGCTG 891
Db 22095 CCGCGCGCGCGCGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 22154
QY 892 ACCACCGAGCCTTCGCGCTCTTCGCGGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCG 951
Db 22155 ACCCG 22214
QY 952 CTCGCGCGCTCCGCTAGGCTCTCCGCGCTCACCACCTAGTACGCGCGCGCTTCG 1011
Db 22215 CCGCGCGCGCGCGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 22274
QY 1012 GTCACGCTGGCAACCGGC 1029
Db 22275 CCG 22292

Search completed: February 23, 2004, 23:23:23
Job time : 132.797 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 19:36:48 ; Search time 552.916 Seconds

(without alignments)
9125.925 Million cell updates/sec

Title: US-09-936-456-1

Perfect score: 1441
Sequence: 1 aaagtaacgctcacagacag.....aaaaaaaaaaaaaaaa 1441

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2308684 seqs, 1750822206 residues

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1435.6	99.6	1474	15	US-10-264-237-1097 Sequence 1097, Ap
2	1413	98.1	1421	14	US-10-187-657-2 Sequence 2, Appli
3	1407.8	97.7	1420	14	US-10-187-657-4 Sequence 4, Appli
4	694.6	48.2	1594	14	US-10-187-657-9 Sequence 9, Appli
5	451.6	31.3	522	14	US-10-187-657-5 Sequence 5, Appli
6	396	27.5	2684	15	US-10-094-749-1195 Sequence 1195, Ap
7	385.8	26.8	450	14	US-10-187-657-7 Sequence 7, Appli
8	354.8	24.6	1029	10	US-09-759-1308-424 Sequence 424, App
9	354.8	24.6	1029	13	US-10-042-431-54 Sequence 54, Appl
10	354.8	24.6	2133	10	US-09-759-1308-423 Sequence 423, App
11	354.8	24.6	2133	13	US-10-042-431-53 Sequence 53, Appl
12	322	22.3	506	14	US-10-187-657-6 Sequence 6, Appli
13	300	20.8	346	14	US-10-187-657-3 Sequence 3, Appli
14	245.8	17.1	2684	15	US-10-094-749-1195 Sequence 1195, Ap
15	200.6	13.9	439	15	US-10-027-632-91970 Sequence 91970, A

16	196.2	13.6	232	9	US-09-783-590-9492 Sequence 9492, Ap
17	170	11.8	250	14	US-10-187-657-8 Sequence 8, Appli
18	64	4.4	512	14	US-10-106-698-3277 Sequence 3277, Ap
19	50.6	3.5	1455	14	US-10-156-761-5889 Sequence 5889, Ap
20	50.6	3.5	9025608	14	US-10-156-761-1 Sequence 1, Appli
21	50	3.5	594	14	US-10-123-155-10 Sequence 10, Appl
22	50	3.5	594	14	US-10-146-731-10 Sequence 10, Appl
23	50	3.5	594	14	US-10-140-472-10 Sequence 10, Appl
24	50	3.5	594	14	US-10-141-761-10 Sequence 10, Appl
25	50	3.5	594	14	US-10-142-885-10 Sequence 10, Appl
26	50	3.5	594	14	US-10-158-790-10 Sequence 10, Appl
27	50	3.5	594	15	US-10-137-871-10 Sequence 10, Appl
28	50	3.5	594	15	US-10-140-923-10 Sequence 10, Appl
29	50	3.5	594	15	US-10-141-756-10 Sequence 10, Appl
30	50	3.5	594	15	US-10-141-759-10 Sequence 10, Appl
31	50	3.5	594	15	US-10-140-805-10 Sequence 10, Appl
32	50	3.5	594	15	US-10-140-864-10 Sequence 10, Appl
33	50	3.5	1362	14	US-10-156-761-7248 Sequence 7248, Ap
34	50	3.5	9025608	14	US-10-156-761-1 Sequence 1, Appli
35	48.4	3.4	511	12	US-10-424-599-56520 Sequence 56520, A
36	48	3.3	234	10	US-09-814-353-4810 Sequence 4810, Ap
37	48	3.3	234	10	US-09-814-353-11107 Sequence 11107, A
38	48	3.3	380	10	US-09-814-353-17491 Sequence 17491, A
39	48	3.3	1035	15	US-10-260-238-272 Sequence 272, App
40	47.6	3.3	277	9	US-09-960-352-12673 Sequence 12673, A
41	47.4	3.3	263	10	US-09-814-353-5246 Sequence 5246, Ap
42	47.4	3.3	263	10	US-09-814-353-11533 Sequence 11533, A
43	47.2	3.3	434	10	US-09-814-353-17917 Sequence 17917, A
44	46.8	3.2	397	14	US-10-198-846-13054 Sequence 13054, A
45	46.6	3.2	332	10	US-09-814-353-17583 Sequence 17583, A

ALIGNMENTS

RESULT 1
US-10-264-237-1097
Sequence 1097, Application US/10264237
Publication No. US20040009491A1
GENERAL INFORMATION:
APPLICANT: Biase et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PA131PI
CURRENT APPLICATION NUMBER: US/10/264,237
CURRENT FILING DATE: 2002-10-04
PRIOR APPLICATION NUMBER: PCT/US01/16450
PRIOR FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: US 60/205,515
PRIOR FILING DATE: 2000-05-19
NUMBER OF SEQ ID NOS: 2876
SOFTWARE: Patentin Ver. 3.1
SEQ ID NO 1097
LENGTH: 1474
TYPE: DNA
ORGANISM: Homo sapiens
US-10-264-237-1097

Query Match	99.6%	Score 1435.6;	DB 15;	Length 1474;
Best Local Similarity	99.8%	Pred. No. 0;		
Matches 1435;	Conservative	2;	Mismatches	1; Indels 0; Gaps 0;
Cy	1	AAAGTAACGGCTACAGACAGTGAAGAAATGATTGCTGCGCCGCTAGAAAACTGTGCG	60	
Db	27	AAAGTAACGGCTACAGACAGTGAAGAAATGATTGCTGCGCCGCTAGAAAACTGTGCG	86	
Cy	61	GTACCAACCCGAGAGCGTTGAGAGCAGCCCACTCCAGCCTTCTTAACGAGAGGTGCA	120	
Db	87	GTACCAACCCGAGAGCGTTGAGAGCAGCCCACTCCAGCCTTCTTAACGAGAGGTGCA	146	
Cy	121	GGAATCAGACTTCAACGAGCCCACTGGTCCAGCCTTGTACGAAAGAGACGCCAAGAC	180	
Db	147	GGAATCAGACTTCAACGAGCCCACTGGTCCAGCCTTGTACGAAAGAGACGTCAAGAC	206	

QY 181 GCGCTCTCCCGCTCCAGGAGAGCCCAAGCTTGTGCTTGGCTTGCCTGCCCCGCTTGCCTGAGC 240
DB 207 GCGCTCTCCCGCTCCAGGAGAGCCCAAGCTTGTGCTTGGCTTGCCTGCCCCGCTTGCCTGAGC 266
QY 241 ACTGGGCGCGGCGAGCATGACCCCTGTGGAACGGCGTACTGCTTTTAAACCCAGAGC 300
DB 267 ACTGGGCGCGGCGAGCATGACCCCTGTGGAACGGCGTACTGCTTTTAAACCCAGAGC 326
QY 301 CGGATGCGCGAGGCTTACAGGCTTCCACTGCTCATGCTTATTTAGTGTTTTGGCTCTA 360
DB 327 CGGATGCGCGAGGCTTACAGGCTTCCACTGCTCATGCTTATTTAGTGTTTTGGCTCTA 386
QY 361 GCAGCAAGCTTCTGCTCATCTTGGCGGAGATCCGTGCGCACTGCGCTGCTTGGTTG 420
DB 387 GCAGCAAGCTTCTGCTCATCTTGGCGGAGATCCGTGCGCACTGCGCTGCTTGGTTG 446
QY 421 GTGAGATTCTTCTAGTCTGTTCATAGCGCGAATAATTGGCTGTGCACTTCACTGCA 480
DB 447 GTGAGATTCTTCTAGTCTGTTCATAGCGCGAATAATTGGCTGTGCACTTCACTGCA 506
QY 481 GAATGTTCTGTGTTACAGTGAACCAACACATCTTACAAAGCTTACAGCGAGCGC 540
DB 507 GAATGTTCTGTGTTACAGTGAACCAACACATCTTACAAAGCTTACAGCGAGCGC 566
QY 541 GTTACAGCCCGTGTGCTGTGCTGTGCGCTGAGGCGCATTAATATTAACCTACAGG 600
DB 567 GTTACAGCCCGTGTGCTGTGCTGTGCGCTGAGGCGCATTAATATTAACCTACAGG 626
QY 601 ACCCAAGTGCATCAGCTGAACGAGACCATTAACCAAGCAGCTGAGCTGCTG 660
DB 627 ACCCAAGTGCATCAGCTGAACGAGACCATTAACCAAGCAGCTGAGCTGCTG 686
QY 661 AAAGAGAATTACCGCGCGAGTACCGCAACGCACTGAGAGGGGCTGCGGAGCCAGTG 720
DB 687 AAAGAGAATTACCGCGCGAGTACCGCAACGCACTGAGAGGGGCTGCGGAGCCAGTG 746
QY 721 CTCTAAGTGGCGGAGAGTTACACACCGAGTACCGCTTGGCGCTGTACCAAGTACAC 780
DB 747 CTCTAAGTGGCGGAGAGTTACACACCGAGTACCGCTTGGCGCTGTACCAAGTACAC 806
QY 781 CTGGCGGAGACATACGCTTCCGCGCAAGCTATGAGGTGGCTTCTGCTTGGCTCTTCC 840
DB 807 CTGGCGGAGACATACGCTTCCGCGCAAGCTATGAGGTGGCTTCTGCTTGGCTCTTCC 866
QY 841 AACGTGCTGCTCTCCACGCGCGCGCTCTACGAGGCTTGGCACTGCTGACCAAGCGA 900
DB 867 AACGTGCTGCTCTCCACGCGCGCGCTCTACGAGGCTTGGCACTGCTGACCAAGCGA 926
QY 901 GCGTTCGCGCTCTTCGGGGGTCTTCGCTTGGCTTCCATCTTACGCTGCGCTTGGCG 960
DB 927 GCGTTCGCGCTCTTCGGGGGTCTTCGCTTGGCTTCCATCTTACGCTGCGCTTGGCG 986
QY 961 CTCGCGCTAGGCTCTCCGCGCTACCACTCAGTACGCGCGCTTCTGGGTCAAGCTG 1020
DB 987 CTCGCGCTAGGCTCTCCGCGCTACCACTCAGTACGCGCGCTTCTGGGTCAAGCTG 1046
QY 1021 GCAACGCGCGTCTGCTGCTCTTCCCTCGAGGGGCGGTGAGTCTCCAGTATGTTGG 1080
DB 1047 GCAACGCGCGTCTGCTGCTCTTCCCTCGAGGGGCGGTGAGTCTCCAGTATGTTGG 1106
QY 1081 CCCAGCGCTCTTCGACCCCTTCTGGAACCAAGCGCAAGGCTGACGCGAGAGAGAGG 1140
DB 1107 CCCAGCGCTCTTCGACCCCTTCTGGAACCAAGCGCAAGGCTGACGCGAGAGAGAGG 1166
QY 1141 GCGTCACTCTTATCTTCGCGAGCCCACTGACAGAGCGGCTCTCCAGACTTAATA 1200
DB 1167 GCGTCACTCTTATCTTCGCGAGCCCACTGACAGAGCGGCTCTCCAGACTTAATA 1226
QY 1201 TGTATCAACCACTAAGCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGAGAT 1260
DB 1227 TGTATCAACCACTAAGCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGAGAT 1286
QY 1261 CGCAGGCGGGAAGAGTGCCCGCGCAAGGCTTGGGCCAGAGAGCTCCAGAGAGGCACTG 1320

DB 1287 CGCAGGCGGGAAGAGAGTGCCCGCGCAAGGCTTGGGCCAGAGAGCTCCAGAGAGGCACTG 1346
QY 1321 AGCGTGTGCGCGAGGAGCTCGGACATCCGAGAGCAGGAGAAAGTCTCTGGGCGA 1380
DB 1347 AGCGTGTGCGCGAGGAGCTCGGACATCCGAGAGCAGGAGAAAGTCTCTGGGCGA 1406
QY 1381 TCTGTAATAAACCTTTTCTTTTGTGTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1438
DB 1407 TCTGTAATAAACCTTTTCTTTTGTGTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1464

RESULT 2

US-10-187-657-2
; Sequence 2, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 1421
; TYPE: DNA
; ORGANISM: Homo sapiens.
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030068311A1 4901066CB1
US-10-187-657-2

Query Match 98.1%; Score 1413; DB 14; Length 1421;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAGTAACGGCTACAGACAGTGAAGAAATAGTTGCTGCGCGGCTAGAAAACTGTGCG 60
DB 9 AAAGTAACGGCTACAGACAGTGAAGAAATAGTTGCTGCGCGGCTAGAAAACTGTGCG 68
QY 61 GTACCAACCCAGAGCGTTGAGAGAGAGCCCACTCCAGGCTTCTTAAAGAGAGGTGCA 120
DB 69 GTACCAACCCAGAGCGTTGAGAGAGAGCCCACTCCAGGCTTCTTAAAGAGAGGTGCA 128
QY 121 GGAATCAAGTTCACAGCCCACTGCTGCTTGTACGCAAGAGAGAGAGAGAGAG 180
DB 129 GGAATCAAGTTCACAGCCCACTGCTGCTTGTACGCAAGAGAGAGAGAGAGAGAG 188
QY 181 GCGCTTCCCGCGTCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
DB 189 GCGCTTCCCGCGTCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 248
QY 241 ACTGGCGCGCGTGCAGCATGACCTGTGGAACGCGCTGCTGCTTGTACCCAGAGCC 300
DB 249 ACTGGCGCGCGTGCAGCATGACCTGTGGAACGCGCTGCTGCTTGTACCCAGAGCC 308
QY 301 CGGATGCGCGAGGCTTTCAGCGTTCCACTGCTCATGCTTATTTAGTGTTTTGGCTCTA 360
DB 309 CGGATGCGCGAGGCTTTCAGCGTTCCACTGCTCATGCTTATTTAGTGTTTTGGCTCTA 368
QY 361 GCAGCAAGTTCCTGCTCATCTTGGCGGGGATCCGTGGCCACTGCGCTGCTTGGTTG 420
DB 369 GCAGCAAGTTCCTGCTCATCTTGGCGGGGATCCGTGGCCACTGCGCTGCTTGGTTG 428
QY 421 GTGAGATTCTTCTAGTCTGTGATAGGCGAGAAATTTGCTGTGCACTTCACTGCA 480

```
Db 429 GTGAGATTCTTCTCAGTCTGTTCATAGGCGCAGAAATTGTGGCTGTGCACCTTCAGTGCA 488
QY 481 GAATGTTTCGTGGGTACAGTGAACACCAACACATCTCAAAAGCCTTCAGCGCAGCGCC 540
Db 489 GAATGTTTCGTGGGTACAGTGAACACCAACACATCTCAAAAGCCTTCAGCGCAGCGCC 548
QY 541 GTTACAGCCCGTGTGGGTCTGTCTGCTGGGGGCTTGAAGGGCATTAATATTACTACAGGG 600
Db 549 GTTACAGCCCGTGTGGGTCTGTCTGCTGGGGGCTTGAAGGGCATTAATATTACTACAGGG 608
QY 601 ACCCAGTGCATCAGCTGAACGAGACCATTTAGTACAACGAGCAGTTCACTGGCGCTCTG 660
Db 609 ACCCAGTGCATCAGCTGAACGAGACCATTTAGTACAACGAGCAGTTCACTGGCGCTCTG 668
QY 661 AAAGAGATTACGCGCGAGTAGCGGAGAACGACCTGGAGAGGGGCTGCCGAGCCAGTG 720
Db 669 AAAGAGATTACGCGCGAGTAGCGGAGAACGACCTGGAGAGGGGCTGCCGAGCCAGTG 728
QY 721 CTCTACCTGGCGAGAGATTACACACGAGTAGCCCTTGCGGCTGTACCAAGTACCACTAC 780
Db 729 CTCTACCTGGCGAGAGATTACACACGAGTAGCCCTTGCGGCTGTACCAAGTACCACTAC 788
QY 781 CTGGCGGACACTACAGCCTCGGCGCAGCTATGGGTGGCGCTTGTGCTTCTGCTCTCC 840
Db 789 CTGGCGGACACTACAGCCTCGGCGCAGCTATGGGTGGCGCTTGTGCTTCTGCTCTCC 848
QY 841 AACGTGCTCTCTCCAGCGCGCGCGCTCTACGAGAGGCTGGCACTGCTGACACCGGA 900
Db 849 AACGTGCTCTCTCCAGCGCGCGCGCTCTACGAGAGGCTGGCACTGCTGACACCGGA 908
QY 901 GCCTTGGCGCTCTTGGGGGTCTTGGCCTTGGCCTCATCTAGCGTGGCGCTTGGCGC 960
Db 909 GCCTTGGCGCTCTTGGGGGTCTTGGCCTTGGCCTCATCTAGCGTGGCGCTTGGCGC 968
QY 961 CTCCGCTTAGGCTCTCCGCGCTACCACTCAGTACGGCGCGCTTCTGGGTACGCTG 1020
Db 969 CTCCGCTTAGGCTCTCCGCGCTACCACTCAGTACGGCGCGCTTCTGGGTACGCTG 1028
QY 1021 GCAACCGGCGCTCTGTGCTCTTCTCTCGAGGGGCGGTGTAGTCTCCAGTATGTTGG 1080
Db 1029 GCAACCGGCGCTCTGTGCTCTTCTCTCGAGGGGCGGTGTAGTCTCCAGTATGTTGG 1088
QY 1081 CCCAGGCTCTTGGCACTCTTCTGACCAAAAGCGCCAAAGACTGCAGCCAGAGAGAGG 1140
Db 1089 CCCAGGCTCTTGGCACTCTTCTGACCAAAAGCGCCAAAGACTGCAGCCAGAGAGAGG 1148
QY 1141 GGCTCAGCTCTTATCTCTCGCGACCACTGCACAGAGGCGCTCTCCAGACTTAA 1200
Db 1149 GGCTCAGCTCTTATCTCTCGCGACCACTGCACAGAGGCGCTCTCCAGACTTAA 1208
QY 1201 TGTATCACCCTAACCCTGTGAGGGGGAACCAATCTGACTCTTCCCGCTTGGACAT 1260
Db 1209 TGTATCACCCTAACCCTGTGAGGGGGAACCAATCTGACTCTTCCCGCTTGGACAT 1268
QY 1261 CGCAGGCGGGAAGAGAGTGGCGCGCAGGCTGGGCGCAGAGAGCTCCAGAGGGCACTG 1320
Db 1269 CGCAGGCGGGAAGAGAGTGGCGCGCAGGCTGGGCGCAGAGAGCTCCAGAGGGCACTG 1328
QY 1321 AGCGTGTGGCGGAGGCTCGACATCCGAGGACCAAGGAAAGTCTCTGGGGCA 1380
Db 1329 AGCGTGTGGCGGAGGCTCGACATCCGAGGACCAAGGAAAGTCTCTGGGGCA 1388
QY 1381 TCTGTAATAAACCTTTTCTTTTGTGTTT 1413
Db 1389 TCTGTAATAAACCTTTTCTTTTGTGTTT 1421
```

RESULT 3
US-10-187-657-4

; Sequence 4, Application US/10187657
; Publication No. US2003006831A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.

```
; APPLICANT: Baughn, Mariah R.  
; APPLICANT: Azimzai, Yalda  
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER  
; FILE REFERENCE: PV-0009 CIP  
; CURRENT APPLICATION NUMBER: US/10/187,657  
; PRIOR APPLICATION NUMBER: PCT/US00/07817  
; PRIOR FILING DATE: 2000-03-22  
; PRIOR APPLICATION NUMBER: 60/139,565  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PERL Program  
; SEQ ID NO 4  
; LENGTH: 1420  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US2003006831A1 3221661CA2  
US-10-187-657-4  
  
Query Match 97.7%; Score 1407.8; DB 14; Length 1420;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1409; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 AAAGTAAAGGCTACAGACAGTGAAGAAATGTTTCGCTCGCGGCTAGAAAACTCTGTG 60  
Db 9 AAAGTAAAGGCTACAGACAGTGAAGAAATGTTTCGCTCGCGGCTAGAAAACTCTGTG 68  
QY 61 GTACCAACCCAGAGCGGTTGAGAGCAGCCCACTCCACGCTTCTTTAAGAGAGTGCA 120  
Db 69 GTACCAACCCAGAGCGGTTGAGAGCAGCCCACTCCACGCTTCTTTAAGAGAGTGCA 128  
QY 121 GAACTCAGACTTCAACAGCCCACTCGGTCCAGCCTTGTATACGAAAGAGAGAGAG 180  
Db 129 GAACTCAGACTTCAACAGCCCACTCGGTCCAGCCTTGTATACGAAAGAGAGAGAG 188  
QY 181 GCGCTCTCCCGCGTCCAGAGCAGCCCACTTGTGCTTGCCTGCGCGCTGCGTGCAGC 240  
Db 189 GCGCTCTCCCGCGTCCAGAGCAGCCCACTTGTGCTTGCCTGCGCGCTGCGTGCAGC 248  
QY 241 ACTCGCGCGGCTGACAGTACCCCTGTGGAACGCGTACTGCTTTTACCAGAGAGCC 300  
Db 249 ACTCGCGCGGCTGACAGTACCCCTGTGGAACGCGTACTGCTTTTACCAGAGAGCC 308  
QY 301 CGGCATCGCGCAGGCTTCAAGCGTTCAGTCTCATGCTTATTTAGTGTGTTGCTCTA 360  
Db 309 CGGCATCGCGCAGGCTTCAAGCGTTCAGTCTCATGCTTATTTAGTGTGTTGCTCTA 368  
QY 361 GCAGCAAGCTTCTGCTCATCTTGCAGGAGATCCGAGCCCACTGCGGCTGTTGGTTG 420  
Db 369 GCAGCAAGCTTCTGCTCATCTTGCAGGAGATCCGAGCCCACTGCGGCTGTTGGTTG 428  
QY 421 GTGAGAGTCTTCTCAGTCTGTTCATAGGCGCAGAAATTGTGGCTGTGACCTTCA 480  
Db 429 GTGAGAGTCTTCTCAGTCTGTTCATAGGCGCAGAAATTGTGGCTGTGACCTTCA 488  
QY 481 GAATGTTTCGTGGGTACAGTGAACACCAACACATCTCAAAAGCCTTCAGCGCAGCGC 540  
Db 489 GAATGTTTCGTGGGTACAGTGAACACCAACACATCTCAAAAGCCTTCAGCGCAGCGC 548  
QY 541 GTTACAGCCCGTGTGGGTCTGTCTGCTGGGGGCTTGAAGGGCATTAATATTACTACAGGG 600  
Db 549 GTTACAGCCCGTGTGGGTCTGTCTGCTGGGGGCTTGAAGGGCATTAATATTACTACAGGG 608  
QY 601 ACCCAGTGCATCAGCTGAACGAGACCATTTAGTACAACGAGCAGTTCACTGGCGCTCTG 660  
Db 609 ACCCAGTGCATCAGCTGAACGAGACCATTTAGTACAACGAGCAGTTCACTGGCGCTCTG 668  
QY 661 AAAGAGATTACGCGCGGAGTAGCGGAGAACGACCTGGAGAGGGGCTGCCGAGCCAGTG 720  
Db 669 AAAGAGATTACGCGCGGAGTAGCGGAGAACGACCTGGAGAGGGGCTGCCGAGCCAGTG 728
```

```
QY 721 CTCTACCTGGCGGAGAAAGTTACACACCGAGTAGCCCTTGCGGCTGTACCAACAGTACAC 780
DB 729 CTCTACCTGGCGGAGAAAGTTACACACCGAGTAGCCCTTGCGGCTGTACCAACAGTACAC 788
QY 781 CTGGCGGGACACTACGCTCTGGGCGACGCTATGAGTGGCGTTCTGCTTCTGCGCTCTCC 840
DB 789 CTGGCGGGACACTACGCTCTGGGCGACGCTATGAGTGGCGTTCTGCTTCTGCGCTCTCC 848
QY 841 AACGTGCTGCTCTCCACGCGCGCGCGCTCTACGAGAGGCTGACCTGTGACCAACCGGA 900
DB 849 AACGTGCTGCTCTCCACGCGCGCGCGCTCTACGAGAGGCTGTGACCTGTGACCAACCGGA 908
QY 901 GCCTTGGCGCTCTTGGGAGGTCTTGGCTTGGCTTCCATCTTACGCTGCGCTCTGCGG 960
DB 909 GCCTTGGCGCTCTTGGGAGGTCTTGGCTTGGCTTCCATCTTACGCTGCGCTCTGCGG 968
QY 961 CTCGCGCTAGGCTCTTCCGCGCTACCACTACGAGAGGCGCGCTTCTGGGTACGCTG 1020
DB 969 CTCGCGCTAGGCTCTTCCGCGCTACCACTACGAGAGGCGCGCTTCTGGGTACGCTG 1028
QY 1021 GCAACCGGCGCTCTGCTGCTCTTCTCTCGAGAGGCGCGTGTGAGTCTCCAGTATGTCG 1080
DB 1029 GCAACCGGCGCTCTGCTGCTCTTCTCTCGAGAGGCGCGTGTGAGTCTCCAGTATGTCG 1088
QY 1081 CCCAGCGCTCTTCCGACCCCTTCTGACCAAAAGCGCAAGAGCTGACGCAAGAGAGAGG 1140
DB 1089 CCCAGCGCTCTTCCGACCCCTTCTGACCAAAAGCGCAAGAGCTGACGCAAGAGAGAGG 1148
QY 1141 GGCTCACTCTTATCTCGCGACCCACTGACAAAGAGCGCGCTCTCCAGACTTAATAA 1200
DB 1149 GGCTCACTCTTATCTCGCGACCCACTGACAAAGAGCGCGCTCTCCAGACTTAATAA 1208
QY 1201 TGTATCACTAATCTGTGAGGAGGAGCCCAATCTGACTCTTCCCGCTTGGGACAT 1260
DB 1209 TGTATCACTAATCTGTGAGGAGGAGCCCAATCTGACTCTTCCCGCTTGGGACAT 1268
QY 1261 CGCAGCGCGGAGAGAGTCCCGCGACGCTGGCGCAGAGAGCTCCAGAAAGGACATG 1320
DB 1269 CGCAGCGCGGAGAGAGTCCCGCGACGCTGGCGCAGAGAGCTCCAGAAAGGACATG 1328
QY 1321 AGCGCTGCTGGCGGAGGCTCGGACATCCGACGAGCAACAGGAAAGTCTCTGGGCGA 1380
DB 1329 AGCGCTGCTGGCGGAGGCTCGGACATCCGACGAGCAACAGGAAAGTCTCTGGGCGA 1388
QY 1381 TCTGTAAATAAACCTTTTCTTTTCTTTT 1411
DB 1389 TCTGTAAATAAACCTTTTCTTTTCTTTT 1419
```

RESULT 4
US-10-187-657-9

```
/ Sequence 9, Application US/10187657
/ Publication No. US20030068311A1
/ GENERAL INFORMATION:
/ APPLICANT: Lasek, Amy K.W.
/ APPLICANT: Baughn, Mariah R.
/ APPLICANT: Azimzai, Yalda
/ TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
/ FILE REFERENCE: PV-0009 CIP
/ CURRENT APPLICATION NUMBER: US/10/187,657
/ CURRENT FILING DATE: 2002-07-01
/ PRIOR APPLICATION NUMBER: PCT/US00/07817
/ PRIOR FILING DATE: 2000-03-22
/ PRIOR APPLICATION NUMBER: 60/139,565
/ PRIOR FILING DATE: 1999-06-16
/ NUMBER OF SEQ ID NOS: 9
/ SOFTWARE: PERL Program
/ SEQ ID NO 9
/ LENGTH: 1594
/ TYPE: DNA
/ ORGANISM: Mus musculus
/ FEATURE:
/ NAME/KEY: misc_feature
```

```
/ OTHER INFORMATION: Incyte ID No. US20030068311A1 110769_Mm.1
US-10-187-657-9
```

Query Match 48.2%; Score 694.6; DB 14; Length 1594;
Best Local Similarity 75.7%; Pred. No. 3.2e-183;
Matches 918; Conservative 0; Mismatches 279; Indels 16; Gaps 4;

```
QY 222 CTGCCGCGCTGCGTGACGACACTCGGCGCGGCTGACGATGACCCGTGTGAACGCGTACT 281
DB 107 CTGCGTGGCTGCGAAGAGTCTCGAATCTTGTGAGGATGACTGCGTGGACGCGGTGCT 166
QY 282 GCCTTTTACCCCGAGCGCGGATGCGGAGGCTTACGCTTCCACTGCTCATCGTTAT 341
DB 167 ACCCTTTTACCCCGAGCGCGGATGCGGAGGCTTACGCTTCCACTGCTCATCGTTAT 226
QY 342 TCTAGTGTGTTTGGCTCTAGACGACGACCTTCTGCTCATCTTGGCGGGATCGTGCCA 401
DB 227 CCGTGTGTTCTTGTCTTGGCTGTAGCTTCTGCTTCTTCTGCTGGAATCGTGCGCA 286
QY 402 CTCGCGCTGTTTGGTGTGAGAGTCTTCTGCTCATCTTGGCGGGATCGTGCCA 461
DB 287 CTCGCGCTGTTTGGTGTGAGAGTCTTCTGCTCATCTTGGCGGGATCGTGCCA 346
QY 462 GGCTGTGACCTTACGAGAGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 521
DB 347 GGCTGTGACCTTACGAGAGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 406
QY 522 AGCCTTACGCGGAGCGCGGTTACAGCCCGTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 581
DB 407 AGCCTTACGCGGAGCGCGGTTACAGCCCGTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 466
QY 582 TAATATTACACTACAGGAGACCCAGTGCATGAGTGAACGAGACCATGACTTACACGA 641
DB 467 TAATATTACACTACAGGAGACCCAGTGCATGAGTGAACGAGACCATGACTTACACGA 526
QY 642 GCAGTTACCTGGCGTGTGAAGAGAAATTACGCGCGGAGTACGCGAAGCACTGAGAA 701
DB 527 GCCTTCACTTGGCGTGTGAAGAGAAATTACGCGCGGAGTACGCGAAGCACTGAGAA 586
QY 702 GGGGCTGCGGAGCCAGTGTCTTCTGCGGAGAAAGTTACACCGAGTAGCCCTTGCGG 761
DB 587 GGGGCTGCGGAGCCAGTGTCTTCTGCGGAGAAAGTTACACCGAGTAGCCCTTGCGG 646
QY 762 CCTGTACACACAGTACACCTGCGGAGACACTAGCCCTCGGCAACGCTATGGTGGCGTT 821
DB 647 GCTGTACACACAGTACACCTGCGGAGACACTAGCCCTCGGCAACGCTATGGTGGCGATT 706
QY 822 CTGCTTCTGCTCTCTTCCAGAGTGTGCTCTTCCAGCGCGCGCGCTCTACGAGGCGCT 881
DB 707 CTGCTTCTGCTCTCTTCCAGAGTGTGCTCTTCCAGCGCGCGCGCTCTACGAGGCGCT 766
QY 882 GGCAGTGTGACACACGAGAGCCTTGTGCGGCTTGTGCGGAGTGTGCGGCTTGTGCGGCT 941
DB 767 GGCAGTGTGACACACGAGAGCCTTGTGCGGCTTGTGCGGAGTGTGCGGCTTGTGCGGCT 826
QY 942 TAGCGTGGCGCTCTGCGCGCTCGGCTAGGCTCTTCCGCGCTGACCACTACAGTACGCGC 1001
DB 827 CAGCGTGGCGCTCTGCGCGCTCGGCTAGGCTCTTCCGCGCTGACCACTACAGTACGCGC 886
QY 1002 CGCCTTCTGAGTACGCTGTGCAACGCGGCTGTGCTGCTTCTTCTGAGGCGCGTGTGT 1061
DB 887 CTCCTTCTGAGTACGCTGTGCAACGCGGCTGTGCTGCTTCTTCTGAGGCGCGTGTGT 946
QY 1062 GAGTCTCCAGTATGTTGCGGCGCGCTGTGCAACGCTTGTGCAACGCGGCGCAAGGA 1121
DB 947 GATTCTCACTATAGTGGCGCGCGCGCTTGTGCTTCTTCTGAGTCTTAAGTGTCAAGA 1006
QY 1122 CT--GCAGCGAGAGAGAGGGGCTTACCTTTATCTCTGCGGCAACCCACTGACAAAGCA 1178
DB 1007 CTGTAGCAACAGGCTTAAGGAATCTTCACTTCACTTCACTTCACTTCACTTCACTTCA 1066
QY 1179 GCGCGCTCTCCAGACTTAATATGTATCAACCACTTATGAGGGGAGCCCAATCTGGA 1238
```

Db 1067 GTTGAAGAGTCCAGACTTAA--TATTACACTCTCCTGTGTAATAATACTGACTCCGGA 1123
QY 1239 CTCCTTCCCCCGCTGGGACATCGAGAGCCGGGAGACAGTCCCGCCAGGCTG-GGCCA 1297
Db 1124 TTCTACCCCTCTTTGGGACCCCATAGACCTGAAGACTGTGTTAAGAGCGCGTGCCAGGA 1183
QY 1298 GGAGAGCTCCAGAGAGGCGACTGAGCGCTGTGGCGCGAGCGCTCGACATCCGACAGCA 1357
Db 1184 GCATAGCTTAGTCTGAGAACTGCTGTCCCTGGCGGACGACAGGGGCGACCCCATGTG 1243
QY 1358 CC-----AGGAAAGTCTCCTGGGCGATCTGTAATAAACCCTTTTCTTTGT 1408
Db 1244 CCTTGCTCTAGAGAAATGGATTCTTCAGAGAACTGTAATAACTTTTGTCTTTT 1303
QY 1409 TTTTAAAAAAA 1421
Db 1304 TTTTCTTCAAAA 1316

RESULT 5

US-10-187-657-5
; Sequence 5, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Jasek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 522
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030068311A1 3221661R6
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 455, 480, 483
; OTHER INFORMATION: a, t, c, g, or other
US-10-187-657-5

Query Match 31.3%; Score 451.6; DB 14; Length 522;
Best Local Similarity 96.7%; Pred. No. 1.3e-115;
Matches 502; Conservative 0; Mismatches 12; Indels 5; Gaps 4;

QY 1 AAAGTAACGGCTACAGACGTGAGAATAAGTTTCGCTCGCCGGCTAGAAAAAAGCTGTGCG 60
Db 9 AAAGTAACGGCTACAGACGTGAGAATAAGTTTCGCTCGCCGGCTAGAAAAAAGCTGTGCG 68
QY 61 GTACCAACCCAGAGCGTTGAGAGAGAGCCCACTCCAGCGCTTCTTAACGAGAGGTGCA 120
Db 69 GTACCAACCCAGAGCGTTGAGAGAGAGCCCACTCCAGCGCTTCTTAACGAGAGGTGCA 128
QY 121 GGAATCAGACTTACCAAGCCCACTCGTCCAGCGCTGTGACGAAAGAGAGCCCAAGAC 180
Db 129 GGAATCAGACTTACCAAGCCCACTCGTCCAGCGCTGTGACGAAAGAGAGCCCAAGAC 188
QY 181 GCGCTCTCCCGCGTCCAGAGAGAGCCCACTGTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db 189 GCGCTCTCCCGCGTCCAGAGAGAGCCCACTGTGCTGCTGCTGCTGCTGCTGCTGCTG 248
QY 241 ACTGGCGCGGCTGAGAGTGAACCTGTGGAACGGCGTACTGCTTTTACCCCAAGCC 300
Db 249 ACTGGCGCGGCTGAGAGTGAACCTGTGGAACGGCGTACTGCTTTTACCCCAAGCC 308

QY 301 CGCATGCCGACAGGCTTTCAGCGCTTCCATGCTCATGCTATTCTAGTGTGCTCTA 360
Db 309 CGGCATGCCGACAGTTCAGAGCGTTCCAGCTGCTCATGCTATTCTAGTGTGCTCTA 367
QY 361 GCAGCAAGCTTCTCTGCTCATCTTGCCCGGAGATCCGTGSCCACTCGCGCTGTTGGTTG 420
Db 368 GCAGCAAG--TTCCTGTCATCTTGCCCGGAGATCCGTGSCCACTCGCGCTGTTGGTTG 425
QY 421 GTGAGAGTCTTCTCAGTCTGTTCATATGCGCGAGAAATTGTGGCTGTGCACTTCACTGCA 480
Db 426 GTGAGAGTCTTCTCAGTCTGTTCATATGNGCAGAAATTGTGGCTGTGCACTTCTGCT-NA 484
QY 481 GAATGTTGTTGGGTACAGTGAACCAACACATCTCTAC 519
Db 485 GAATGTTGTTGGGTACAGTGAACCAACATCTCTAC 522

RESULT 6

US-10-094-749-1195/c
; Sequence 1195, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUKIO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKI, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1195
; LENGTH: 2684
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-1195

Query Match 27.5%; Score 396; DB 15; Length 2684;
Best Local Similarity 97.6%; Pred. No. 8.8e-100;
Matches 402; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1007 TCTGGGTCAAGCTGGCAACCGCGCTCCTGTGCTCTTCTCTCGAGGGCGCTGTGACTC 1066
Db 2363 TTTGATCCCAACCGCAAGCGCTCCTGTGCTCTTCTCTCGAGGGCGCTGTGACTC 2304
QY 1067 TCCAGTATGTTGGGCCAGCGCTCTTCCAGCCCTTCTGACCAAGCGCCCAAGACTGCA 1126
Db 2303 TCCAGTATGTTGGGCCAGCGCTCTTCCAGCCCTTCTGACCAAGCGCCCAAGACTGCA 2244
QY 1127 GCCAGAGAGAGGGGGCTCACTCTTATCTCTGCGGACCACTGACAGCAGCGCTC 1186
Db 2243 GCCAGAGAGAGGGGGCTCACTCTTATCTCTGCGGACCACTGACAGCAGCGCTC 2184

QY	1187	TCCCGAGCTTAAATGTATCACCCTAACCTGTGAGGGGACCCCAATCTGGACTCTTCC	1246
Db	2183	TCCCGAGCTTAAATGTATCACCCTAACCTGTGAGGGGACCCCAATCTGGACTCTTCC	2124
QY	1247	CCGCTTGGGACATCGCAGGCCGGGAAAGCAGTGTCCCGCCAGGCTTGGGCGAGAGAGCTC	1306
Db	2123	CCGCTTGGGACATCGCAGGCCGGGAAAGCAGTGTCCCGCCAGGCTTGGGCGAGAGAGCTC	2064
QY	1307	CAGGAAGGGCACTGAGCGCTGTGGCGCGAGGCTTCGCAATCCGCAAGCACCAGGAAA	1366
Db	2063	CAGGAAGGGCACTGAGCGCTGTGGCGCGAGGCTTCGCAATCCGCAAGCACCAGGAAA	2004
QY	1367	GTCTCCTGGGGCGATCTGTAAATAAACCTTTTTTCTTTTGTTTTTTAAAAA	1418
Db	2003	GTCTCCTGGGGCGATCTGTAAATAAACCTTTTTTCTTTTGTTTTTTAAAAA	1952

RESULT 7

```

US-10-187-657-7
; Sequence 7, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzal, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 450
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID NO. US20030068311A1 SEQID4510D1
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 140, 165, 203, 237, 269, 289, 307, 338, 344, 347, 354, 390, 430, 445
; OTHER INFORMATION: a, t, c, g, or other
US-10-187-657-7

```

Query Match 26.8%; Score 385.8; DB 14; Length 450;

Best Local Similarity 93.7%; Pred. No. 2,7e-97;
Matches 404; Conservative 0; Mismatches 26; Indels 1; Gaps 1;

QY	616	CTGAACGAGAGACCATTGTACTACAA	CGAGCAGTTCACCTGGCGGTCTGAAGAAGAAATTACGCC	675
Db	21	CTGAACGAGAGACCATTGTACTACAA	CGAGCAGTTCACCTGGCGGTCTGAAGAAGAAATTACGCC	80
QY	676	GCGGAGTACGCGAACGCACCTGGAGAAAGGGGCTGCCGGACCCAGTGTCTTA	CTCTGCGGAG	735
Db	81	GCGGAGTACGCGAACGCACCTGGAGAAAGGGGCTGCCGGACCCAGTGTCTTA	CTCTGCGGAG	140
QY	736	AAGTTCACACCGAGTAGACCCCTTGGCGGCTGTATCCACCAAGTATCCACCTGGCGGGACATAC		795
Db	141	AAGTTCACACCGAGTAGACCCCTTGGCGGCTGTATCCACCAAGTATCCACCTGGCGGGACATAC		200
QY	796	GCCTCGGCCACCGCTATGGGTGGCGTTCGTCTTCTGGCTCCTCTCCAACGHTGCTCTCC		855
Db	201	GCNTCGGCCACCGCTATGGGTGGCGTTCGTCTTCTGGNTCCTCTCCAACGHTGCTCTCC		260
QY	856	ACGCGGGCCCCCGCTCTACCGAGGGCTGGCAGCTGTACCAACCGGAGGCTTTCGGCTCTTC		915
Db	261	ACGCGGGCCCCCGCTCTACCGAGGGCTGGNACTGTCTGACCAACCGGAGGCTTTCGGCTCTTC		320
QY	916	GGGCTCTTCGCTTGGCCTCCACTCTAGCGGTGCCGCTCTGCCCCGCTCCGCTTAGGCTCC		975

Db 321 GGGGTCTTGCCNTGCCAATTCNAGGTGNCGCTTGCCAGATACGCTAGGCTCC 380

QY 976 TCCGCGTCAACCACTAGTACGGCGCGCTTCTGGGTACGCTGGCAACGGCGTCTCG 1039

Db 381 TCCGCG-TCANCAATCAGTAACGGAACGGCTTCTGGGTCAAGCTGGCAANAGCGTCTCG 439

QY 1036 TGCCTCTTCT 1046

Db 440 TGCCTNTTCT 450

RESULT 8

```

US-09-759-130B-424
; Sequence 424, Application US/09759130B
; Publication No. US20030022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: Kirst, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodheartl, Andrew
; APPLICANT: Holtzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; TITLE OF INVENTION: USES.
; FILE REFERENCE: MP100-5350NMIM
; CURRENT APPLICATION NUMBER: US/09/759,130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 424
; LENGTH: 1029
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-759-130B-424

```

Query Match 24.68; Score 354.8; DB 10; Length 1029;

Best Local Similarity 64.5%; Pred. No. 1.8e-88;
Matches 546; Conservative 0; Mismatches 297; Indels 3; Gaps 1;

QY 279 ACTGCTTTTAAACCCAGCCCCGCGATGCCGCGCTTCAGCGTTCCACTGCTCATCGT 338
Db 21 ATTCCCCCTCATGCTGGGCCCAAGCAACCTTCCGATGACACCACTTTGGCCAGCAT 80
QY 339 TATTCTAGTGTTTTGGCTCTAGACGAAGCTTCCGTCTCATCTTGGCCGGGATCCGTGG 398


```

Db      81 CATCATGATCTTTCTGACTGCACTGGCCACGTTTCATCTGCTGCGCATTCGGGG 140
QY      399 CCACTCGCGCTGGTTTGGTTGGTGAGAGTTCTTCTCAGTCTGTTCAIAGCGCAGAAAT 458
Db      141 AAAGACGAGGCTGTTCTGGCTGCTTCCGGTGGTGACCAGCTTATTCACTGGGGCTGCAT 200
QY      459 TGTGCTGTGCACTTCACTGACGAATGGTTCGTGGGTACAGTGAACACCAACATCTCTA 518
Db      201 CCTGGCTGTGAATTTCACTTCTGAGTGTCTGTGGCCAGGTACAGACCAACATCATATA 260
QY      519 CAAAGCCTTCAGCGCGACCGCGCTTACAGCCCGTCTGCTCTGCTGGGCTTGAGGG 578
Db      261 CAAGGCTTCAGTTCTGAGTGAATCAGCGCTGATATTGGGCTGCAGTCCGGCTGGGTG 320
QY      579 CATTAATATTACACTCACAGGAGCCCGAGTGCATCAGTGAACGAGACCATTTGACTACA 638
Db      321 AGTCAACATCACACTCACAGGAGCCCGCTGACAGCTGAATGAGACCATCAATTACAA 380
QY      639 CGAGCACTTCACTGCGGCTTGAAGAAGATTACGCCCGGAGTACGCGAAGCACTGGA 698
Db      381 CGAGGAGTTCACTGCGGCTTGGTGAAGACTATGCTGAGAGTGTGCAAGGCTCTGGA 440
QY      699 GAAGGGGCTGCGGACCCAGTCTCTACCTGCGGAGAAATTACACCGAGTACCCCTTG 758
Db      441 GAAGGGGCTGCGGACCCAGTCTCTACCTGCGGAGAAATTACACCGAGTACCCCTTG 500
QY      759 CGGCTGTACACCAAGTACACCACTGCGGAGAAATTACCGCTGCGGAGTATGGTGGC 818
Db      501 TGGCTTATACCGCCAGTACCGCTGCGGAGAAATTACCGCTGCGGAGTATGGTGGC 560
QY      819 GTTCTGTCTGCTGCTCTCTCCAAAGTGTCTCTCCACGCGCGCCCGCTCTACGAGG 878
Db      561 ATTCTGTCTGCTGCTGCTGCGCAATGTGTCTCTCCATGCTGTGTATATGTGG 620
QY      879 CCTGCACTGTGACCAACCGGAGCTTTCGCGCTCTGCG--GGGTCTTCCGCTTGGCCTC 935
Db      621 CTACATGTATTGGCCACGCGCATTTCCAGCTGTGTGGCTCTGCTCTTCTTCCATGCG 680
QY      936 CATCTCTAGCGTGCCTCTGCGCGCTCTGCGCTGAGCTCTCCGCGCTCACCACTCAGTA 995
Db      681 CACATCACTCACTCACTCACTGCTCTGCGCGCTCTGCGCGCTCTGCGCTCACCACTCAGTA 740
QY      996 CGGCGCGCTTCTGCGTCACTGCGGAGAAACCGGCTCTGCTCTTCTTCCGAGGGGC 1055
Db      741 TGGGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 800
QY      1056 CGTGTGAGTCTCCAGTATGTTGCGGCCAGCGCTCTTTCGACCCCTTCTGACCAAGCGC 1115
Db      801 TATGGCGGTGGCCACAGAGATGAGCCTTCAAGGCTTCTTCAACAGAGTGT 860
QY      1116 CAAGGA 1121
Db      861 GGATGA 866

```

RESULT 9

```

US-10-042-431-54
; Sequence 54, Application US/10042431
; Publication No. US20020182675A1
; GENERAL INFORMATION:
; APPLICANT: MCCARTHY, Sean A
; APPLICANT: BARNES, Thomas M
; APPLICANT: FRASER, Christopher C
; APPLICANT: SHARP, John D
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC,
; FILE REFERENCE: 10:47-6U2
; CURRENT APPLICATION NUMBER: US/10/042,431
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24

```

```

; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 54
; LENGTH: 1029
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-042-431-54

```

```

Query Match      24.6%; Score 354.8; DB 13; Length 1029;
Best Local Similarity 64.5%; Pred. No. 1.8e-88;
Matches 546; Conservative 0; Mismatches 297; Indels 3; Gaps 1;

```

```

QY      279 ACTGCCCTTTTACCCCGAGCCCGGATGCGGAGCTTCAAGCTTCACTGCTCATCGT 338
Db      21 ATCCCTCTATGCTGGCCCGCAAGCCCACTTCCGATGACACCACTTTGGCAGCAT 80
QY      339 TATTCTAGTGTGTTGGCTCTAGCAGCAAGCTTCTGCTCATCTTCCGGGATCCGTGG 398
Db      81 CATCATGATCTTCTGACTGCACTGGCCACGTTCAATCTCATCTGCTGCGCATTCGGGG 140
QY      399 CCACTGCGGCTGTTTGGTGTGAGAGTCTTCTCAGTCTGTTCAATAGCGCAGAAAT 458
Db      141 AAAGACGAGGCTGTTCTGCGCTGCTTCCGGTGGTGACCAAGTATTATCATCGGGCTGCAAT 200
QY      459 TGTGCTGTGCACTTCACTGACGAATGTTCTGTGGTACAGTGAACCAACACATCTCTA 518
Db      201 CCTGCTGTAAATTTCACTTCTGAGTGTCTGTGGCCAGGTACAGCAACACATCATATA 260
QY      519 CAAAGCCTTCAGCGCAGCGCGCTTACAGCCCGTCTGCTGCTGCTGCTGCTGCTGCTG 578
Db      261 CAAGGCTTCACTTCTGAGTGTGATCAGCGCTGATATTGGGCTGCAAGTCTGGGCTGGGTG 320
QY      579 CATTAATATTACACTCACAGGAGCCCGAGTGCATCAGCTGAACGAGACCATTTGACTACAA 638
Db      321 AGTCAACATCACACTCACAGGAGCCCGCTGACAGCTGAATGAGACCATCAATTACAA 380
QY      639 CGAGCACTTCACTGCGGCTTGAAGAAGATTACCGCGCGGAGTACGCGAAGCACTGGA 698
Db      381 CGAGGAGTTCACTGCGGCTTGGTGAAGACTATGCTGAAGAGTGTGCAAGGCTCTGGA 440
QY      699 GAAGGGGCTGCGGACCCAGTGTCTTACCTGCGGAGAAATTACACCGAGTACCCCTTG 758
Db      441 GAAGGGGCTGCGGACCCAGTGTGTGTGTACTGAGAAATTCACTCAAGAAAGCCCATG 500
QY      759 CGGCTGTACCAACAGTACCACTGCGGAGAACTACGCTTCCGCGCTATAGGGTGGC 818
Db      501 TGGCTTATACCGCCAGTACCGCTTCCGCGGAGAACTACGCTTCCGCGCTATAGGGTGGC 560
QY      819 GTTCTGCTTGTGCTCTCTTCCAAAGTGTCTCTCAAGCGCGCGCCCGCTCTACGGAAG 878
Db      561 ATTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 620
QY      879 CTTGGCACTGCTGACCAACCGGAGCTTGGCGCTCTGCG--GGGTCTTCCGCTTGGCCTC 935
Db      621 CTACATGTATTGGCCACGCGCATTTCCAGCTGTGTGCTGTCTTCTTCCATGCGC 680
QY      936 CATCTCTAGCGTGCCTCTGCGCGCTCTGCGCTAGGCTCTTCCGCGCTCACCACTCAGTA 995
Db      681 CACATCACTCACTCACTCACTGCTCTGCGCGCTCTGCGCGCTCTGCTGCTCATACTCA 740
QY      996 CGGCGCGCTTCTGCGTCACTGCGGAGAAACCGGCTCTGCTGCTCTTCTTCCGAGGGGC 1055
Db      741 TGGGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 800
QY      1056 CGTGTGAGTCTCCAGTATGTTGCGGCCAGCGCTCTTTCGCAACCCCTTCTGAGCAAGCGC 1115
Db      801 TATGGCGGTGGCCACAGAGATGAGCCTTCAAGGCTTCTTCAACAGAGTGT 860
QY      1116 CAAGGA 1121
Db      861 GGATGA 866

```



```
RESULT 10
US-09-759-130B-423
; Sequence 423, Application US/09759130B
; Publication No. US20030022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: KIRST, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wighton, Nicolas
; APPLICANT: Goodearl, Andrew
; APPLICANT: Holtzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; TITLE OF INVENTION: USES.
; FILE REFERENCE: MPIO-535OMNIM
; CURRENT APPLICATION NUMBER: US/09/759,130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 423
; LENGTH: 2133
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-759-130B-423

Query Match      24.6%; Score 354.8; DB 10; Length 2133;
Best Local Similarity 63.8%; Pred. No. 2.5e-88;
Matches 555; Conservative 0; Mismatches 312; Indels 3; Gaps 1;
```

```
Db 306 CCAGTCAGCACCAACATCATACAGGCTTCAGTTCTGAGTGATGACGCGTGATAT 365
QY 555 CGGTCTGCTCGTGGGCTTGAGGCGATTATATTTACACTCACAGGACCCCACTGATCA 614
Db 366 TGGGCTGCAAGTGGGCTGGGTGAGTCAACATCACACTCACAGGACCCCGTGACAGA 425
QY 615 GCTGACGAGACCATTTGACTACACGAGCAGTTCACTGGCGTCTGAAGAATTTACGC 674
Db 426 GCTGAATGAGACCATCAATTACACGAGAGTTCACTGGCGCTGGGTGAACTATGC 485
QY 675 CGCGAGTACGCGAACGCACTGGAGAGAGGCTGCCGACCAAGTGTCTACCTGGCGGA 734
Db 486 TGAGAGTGTGCAAGGCTCTGAGAGAGGCTGCCAGACCTGTGTGTAACCTAGCTGA 545
QY 735 GAAGTTCACACCGAGTAGCCCTTGCGGCTGTACCAACAGTACCACTGGCGGACACTA 794
Db 546 GAAGTTCACCAAGAACCCATGTGGCTTATACCGCAGTACCGCTGGCGGACACTA 605
QY 795 CGCCTCGGCAAGCTATAGGCTGGCTTCTGCTTCTGCTCTCCCAACGCTGCTCTC 854
Db 606 CACCTGAGCCATGATAGGCTGGCAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTC 665
QY 855 CACGCGGCGCCGCTCTACGAGGCTTGCACTGCTGACCAACCGAGCCTTGCGCTCT 914
Db 666 CATGCTGTGCTGATATATGTGGCTACATGCTATTTGGCCACGGGCATCTTCAAGCTGT 725
QY 915 CG--GGGCTTTCGCTTGCCCTTCATCTTACGCTGCGCTGCGCTGCGCTGCGCTAGG 971
Db 726 GGCTGTGCTTCTTCTCCATGCGCAATCACTCACTCACTCACTCACTCACTCACTGCG 785
QY 972 CTCCTCGGCGCTACACACTCAGTACGCGCGCGCTTCTGGGTCAAGCTGGCAACGGCGCT 1031
Db 786 CGCTTGTGTGCTGATCACTACCACTGCGCGCTGCTTCTGATCACTTGAACACAGGACT 845
QY 1032 CCTGTGCTTCTCTGAGGAGGCGCGCTGTGAGTCTCCAGTATGTTGGGCCAGCGCTCT 1091
Db 846 GCTGTGTGTGCTGCTGCGCGCTGCTATGTGCGGTGCGCCACAGAGATGACGCTCAAGGCT 905
QY 1092 TCGCACCTTCTTGACCAAGCGCCAAAGA 1121
Db 906 GAAGCTTCTTCAACCAAGAGTGTGATGA 935
```

```
RESULT 11
US-10-042-431-53
; Sequence 53, Application US/10042431
; Publication No. US20020182675A1
; GENERAL INFORMATION:
; APPLICANT: MCCARTHY, Sean A
; APPLICANT: BARNES, Thomas M
; APPLICANT: FRASER, Christopher C
; APPLICANT: SHARP, John D
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC,
; TITLE OF INVENTION: PREVENTIVE, THERAPEUTIC, AND OTHER USES
; FILE REFERENCE: 10147-6U2
; CURRENT APPLICATION NUMBER: US/10/042,431
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53
; LENGTH: 2133
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-042-431-53

Query Match      24.6%; Score 354.8; DB 13; Length 2133;
Best Local Similarity 63.8%; Pred. No. 2.5e-88;
Matches 555; Conservative 0; Mismatches 312; Indels 3; Gaps 1;
```

```

QY 255 CAGCATGACCCCTGTGGAACGGGCTACTGCTTTTACCCCGGATGCCGAGG 314
Db 66 CAAGATGCTACTTTGGGACACACATTCCTTCTATGCTGGCCCCAAGCCACTTCCC 125
QY 315 CTTACGCTTCCACTGCTCATGCTATTTCTAGTGTGCTTTTGGCTTAGACGAGCTTCC 374
Db 126 GATGACACCACTTTGGGACGATCATGATCTTTCTGACTGACTGGCCAGCTTCAT 185
QY 375 GCTCATCTTGCCGGGATCCGCTGGGCACTCGGCTGCTTTGTTGGTGAAGTCTTCT 434
Db 186 CGTCATCCTGCTGCAATTCGGGAAAGACGAGGCTGCTTCTGGCTGCTTGGGTGTGAC 245
QY 435 CAGTCTGTATAGGCGGAGAAATGTGGCTGTGCACTTCAAGTGAAGATGTTCTGGG 494
Db 246 CAGCTTATTCATCGGGCTGCAATCCTGGCTGTGAATTTCAATTCAGTGTGAGTGTGGG 305
QY 495 TACAGTGAACCAACACATCTCTACAAGCCTTCAGCGCAGCGCGCTTACAGCCGCTGT 554
Db 306 CCAGGTACGACCAACACATCATACAGGCTTCAAGTTCAGTGTGATCAGCGCTGATAT 365
QY 555 CGGTCTGCTCGTGGGCTTGAGGGGCAATATATATACCTACAGAGGAGCCGAGTGCATCA 614
Db 366 TGGGCTGAGGTGCGGCTGGGTGAGTCAACATCACTACAGGAGACCCCGTGCAGCA 425
QY 615 GCTGAACGAGACCATGATGACTACACAGAGCAATCACTGCGCTGGAAGAGATTAAGC 674
Db 426 GCTGAATGAGACCATCAATTAACAAGAGAGTCACTGCGCTGGGTGAGAACTATGC 485
QY 675 CGCGGAGTACGCGAAGCACTGAGAGAGGGGCTGCCGAGCCAGTGTCTTACCTGGCGGA 734
Db 486 TGAGGAGTGTGCAAGAGCTCTGAGAGAGGGGCTGCCAGACCTGTGTGACTAGCTGA 545
QY 735 GAAGTTACACCGAGTACCCCTTGCGGCTGTACCAACAGTACCACTGGCGGAGACATA 794
Db 546 GAAGTTCACTCCAGAGGCCCATGTGGCTATACCGCCAGTACCGGCTGGCGGAGACATA 605
QY 795 CGCCTCGGCCACGCTATGGGTGGGCTTCTGCTTCTGCTTCTTCCAAAGTGTGCTTC 854
Db 606 CACCTCAGCCATGCTATGGGTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 665
QY 855 CACGCCGCGCGCTCTACGAGAGGCGCTGGCACTGCTGACCAACCGGAGCCTTGGCGCTCT 914
Db 666 CATGCTGTGCTGTATATGTGTGGCTACATGCTATTTGGCCACGGGCACTTCCAGCTGTT 725
QY 915 CG--GGGTCTTGGCCTTGGGCTTCCATCTCTAGCGTGGCGCTGTGCGGCTTGG 971
Db 726 GGCTCTGCTTCTTCTTCCATGGCCACATCACTCACTCACTCACTCACTCACTCACTCACT 785
QY 972 CTCCTCCGCGCTACCACTCAGTACGCGCGCGCTTCTGGGTACAGCTGGCAACCGGCT 1031
Db 786 CGCTTCTGTGCTCATATCAACCATGGGCTTCTGATCATATGACCAAGAGACT 845
QY 1032 CCTGTGCTTCTTCTCGGAGGGGCGGCTGTGAGTCTCCAGTATGTTGGGCCAGGCTCT 1091
Db 846 GCTGTGTGTGCTGCTGGGCTGTGCTATGGCGGCTGGCCACAGGATGAGCCTCAAGGCT 905
QY 1092 TCGCACCCCTTCTGACCAAGCGCCAGGA 1121
Db 906 GAAGGCTTCTTCAACCAAGAGTGTGATGA 935

```

```

RESULT 12
US-10-187-657-6
; Sequence 6, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; CURRENT FILING DATE: 2002-07-01

```

```

; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 6
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030068311A1 SBQA03652D1
; NAME/KEY: unsure
; LOCATION: 87, 89, 93, 95, 124, 178, 227, 253, 262, 279, 285, 287, 293-294, 298,
; LOCATION: 300, 303, 312, 319-321, 323, 325, 327, 332-333, 336, 343-344, 349, 377,
; LOCATION: 405, 409-410, 412, 415, 420, 423, 429, 432, 436-437, 444-445, 448, 452,
; LOCATION: 466, 469-471, 479, 483, 485, 492, 494, 499, 501, 504
; OTHER INFORMATION: a, t, c, g, or other
US-10-187-657-6

```

```

Query Match      22.3%; Score 322; DB 14; Length 506;
Best Local Similarity 86.3%; Pred. No. 1.8e-79;
Matches 353; Conservative 0; Mismatches 54; Indels 2; Gaps 2;

```

```

QY 464 CTGTGACCTTCAAGTGCAGATGTTGCTGGGTACAGTGAACCAACACATCTTACAAG 523
Db 17 CTGTGACCTTCAAGT-CAGATGTTGCTGGGTACAGTGAACCAACACATCTTACAAG 75
QY 524 CTTTCAGCGCAGCGCGCTTACAGCCCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 583
Db 76 CTTTCAGCGCAGCGCGCTTACAGCCCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 135
QY 584 ATATTACACTCAGAGGAGGAGCCAGTGCATCAGCTGAACGAGACCATTAACAAGAGC 643
Db 136 ATATTACACTCAGAGGAGGAGCCAGTGCATCAGCTGAACGAGACCATTAACAAGAGC 195
QY 644 AGTTACCTGGGCTCTGAAGAAGATTACGCCGCGGAGTACCGGAGCACTGGAGAAAG 703
Db 196 AGTTACCTGGGCTCTGAAGAAGATTACGCCGCGGAGTACCGGAGCACTGGAGAAAG 255
QY 704 GGTGCGCGGAGCCAGTGTCTTACCTTGGCGGAGAGTTCACACCGAGTACCCCTTGGCGCC 763
Db 256 GGTGCGCGGAGCCAGTGTCTTACCTTGGCGGAGAGTTCACACCGAGTACCCCTTGGCGCC 315
QY 764 TGTACCAACAGTACCACTTGGCGGAGCACTACGCTTGGCCACGCTATGGGTGGCTTCT 823
Db 316 TGTNNNNANNGNACCAANTNCGGAGCANNATACANTCGGCCACGCTATGGGTGGCTTCT 375
QY 824 GCTTCTGGCTCC-TCTCCAAGTGTGCTCTCCACGCGCGCGCGCTCT 871
Db 376 GNTTCTGGCTCTTNTCCAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 424

```

```

RESULT 13
US-10-187-657-3
; Sequence 3, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program

```

```
; SEQ ID NO 3
; LENGTH: 346
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US2003068311A1 1752794F6
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 91, 188, 206, 291, 337
; OTHER INFORMATION: a, t, c, g, or other
US-10-187-657-3
```

```
Query Match      20.8%; Score 300; DB 14; Length 346;
Best Local Similarity 94.6%; Pred. No. 2.1e-73;
Matches 330; Conservative 0; Mismatches 15; Indels 4; Gaps 2;
```

```
QY 927 CTTGGCCTCCATCTCTAGCGGTGCGGCTCTGCCGCTCCGCTAGGCTCCTCCGCGCTAC 986
    |||||
Db 1 CTTGGCCTCCATCTCTAGCGGTGCGGCTCTGCCGCTCCGCTAGGCTCCTCCGCGCTAC 60

QY 987 CACTCAGTACGGCGCGGCTCTCTGGGTCAACGCTGGCAACCGGCTCTGCTCTTCT 1046
    |||||
Db 61 CACTCAGTACGGCGCGGCTCTCTGGGTCAACGCTGGCAACCGGCTCTGCTCTTCT 120

QY 1047 CGAGGGGGCGGTGTAGTCTCCAGTATGTTCGGCCCGGCTCTTCCGACCCCTTCTGA 1106
    |||||
Db 121 CGAGGGGGCGGTGTAGTCTCCAGTATGTTCGGCCCGGCTCTTCCGACCCCTTCTGA 180

QY 1107 CCAAGGGCCAGAGACTGCAGCCAGAGAGAGAGGGGGCTCACCTCTTATCCTCGGCA-CC 1165
    |||||
Db 181 CCAAGGGCCAGAGACTGCAGCCAGAGAGAGAGGGGGCTCACCTCTTATCCTCGGCA-CC 240

QY 1166 CACTGCAAGAGAGCGGCTCTCCGCAAGTAAATGTATCACTAAGCTGTAGGGG 1225
    |||||
Db 241 CACTGCAAGAGAGCGGCTCTCCGCAAGTAAATGTATCACTAAGCTGTAGGGG 300

QY 1226 GACCCAATCTGACTCTTCCCGGCTTGGGACATGCAGGCGGGAG 1274
    |||||
Db 301 GACCCAATCTGACTCTT--CCCGGCTTGGGACATGCAGTCCGGAAG 346
```

RESULT 14

US-10-094-749-1195

Sequence 1195, Application US/10094749

Publication No. US20030219741A1

GENERAL INFORMATION:

APPLICANT: ISOGAI, TAKAO

APPLICANT: SUGIYAMA, TOMOYASU

APPLICANT: OTSUKI, TETSUJI

APPLICANT: WAKAMATSU, AI

APPLICANT: SATO, HIROYUKI

APPLICANT: ISHII, SHIZUKO

APPLICANT: YAMAMOTO, JUN-ICHI

APPLICANT: ISONO, YUTKO

APPLICANT: HIO, YURI

APPLICANT: OTSUKA, KAORU

APPLICANT: NAGAI, KEIICHI

APPLICANT: IRIE, RYOTARO

APPLICANT: TAMECHIKA, ICHIRO

APPLICANT: SEKI, NAOHICO

APPLICANT: YOSHIKAWA, TSUTOMU

APPLICANT: OTSUKA, MOTOYUKI

APPLICANT: NAGAHARA, KENJI

APPLICANT: MASUHO, YASUHIKO

TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA

FILE REFERENCE: 084335/0160

CURRENT APPLICATION NUMBER: US/10/094,749

PRIOR FILING DATE: 2002-03-12

PRIOR APPLICATION NUMBER: 60/350,435

PRIOR FILING DATE: 2002-01-24

PRIOR APPLICATION NUMBER: JP 2001-328381

PRIOR FILING DATE: 2001-09-14

```
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1195
; LENGTH: 2684
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-1195
```

```
Query Match      17.1%; Score 245.8; DB 15; Length 2684;
Best Local Similarity 67.4%; Pred. No. 7.3e-58;
Matches 362; Conservative 0; Mismatches 172; Indels 3; Gaps 1;
```

```
QY 588 TACACTCAGAGAGACCCAGTGCATCAGTGAACGAGACCATTAAGTACACGAGAGATT 647
    |||||
Db 250 TGCAATCTGGGAGACCCCGGTGACGAGCTGAATGAGACCATTAATTAACGAGAGATT 309

QY 648 CACCTGGGCTGTGAAGAAGATTACGCCGCGAGTACGGGAACGCACTGAGAGAGGGGCT 707
    |||||
Db 310 CACCTGGGCTGTGGTGAAGACTATGCTGAGAGATTGCAAGAGCTCTGAGAGAGGGGCT 369

QY 708 GCCGAGCCAGTGTCTTACCTGGCGGAGAGTTCAACCGAGTAGCCCTTGGGCGCTGTA 767
    |||||
Db 370 GCCGAGCCAGTGTGTGTACCTAGCTGAGAGAGTTCACTCAAGAGAGCCCATGTGGCTATA 429

QY 768 CCACCAAGTACCACTGGCGGAGACATACGCTCGGCCACGCTATGGGTGGCGTCTGCTT 827
    |||||
Db 430 CCGCCAGTACCGCTGTGGCGGAGACATACACCTCAGCCATGTCTATGGGTGGCATTCCTTG 489

QY 828 CTGGCTCTCTCCAAAGTGTCTGTCTTCCACGCGCGGCCCGCTCTACGAGGCGCTGGCACT 887
    |||||
Db 490 CTGGCTGTGGCCAAATGTATGTCTTCCATGCTGTGTGTATGTGTGTATGTGTGTATGTGT 549

QY 888 GCTGACCAACGAGAGCTTCCGCGCTCTG--GGTCTTGGCTTGGCGCTCAATCTTAG 944
    |||||
Db 550 ATTGGCCAGGGGATCTTCCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 609

QY 945 CGTGGCGCTGTGGCGGCTTCCGCTAGGCTCTCCGCGCTACCACTAGTACGAGCGCGC 1004
    |||||
Db 610 CACTCAGCTGTCTCCGCTGACCTGTGGCGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 669

QY 1005 CTTCTGGGTCAAGCTGGCAACGCGGCTGTGCTCTTCTCGAGGGGCGGTGTAG 1064
    |||||
Db 670 CTTCTGGATCAATTGACCAAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 729

QY 1065 TCTCCAGTATGTTCGGCCAGCGGCTTTCGACCTTCTGACCAAGCGCCAGGA 1121
    |||||
Db 730 GGCCCAAGAGTACAGCTTCAAGGCTTCTTCAACCAAGAGTGTGTAGTA 786
```

RESULT 15

US-10-027-632-91970

Sequence 91970, Application US/10027632

Publication No. US20030204075A9

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

POLYMORPHISMS IN THE HUMAN GENOME

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 91970
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(439)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-91970

Query Match 13.9%; Score 200.6; DB 15; Length 439;
Best Local Similarity 97.3%; Pred. No. 1.3e-45;
Matches 214; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 812 GGGTGGCGTTCTGCTTCTGGCTCCTCTCCACAGTGTGCTCTCCACGCGCGCGCTCT 871
Db 133 GGGTGGCGTTCTGCTTCTGGCTCCTCTCCACAGTGTGCTCTCCACGCGCGCGCTCT 192

QY 872 ACGAGGCGCTGGCACTGCTG-ACCACCGGAGCCTTGGCGCTCTTGGGGTCTTGGCTTG 930
Db 193 GCGGAGCGCTGGCNCCTGCTGNACCAACCGAGCCTTGGCGCTCTTGGGGTCTTGGCTTG 252

QY 931 GCCTCCATCTCTAGCGTGCGCTCTGCGCGCTCCGCGCTAGGCTCCTCCGCGCTCACT 990
Db 253 GCCTCCATCTCTAGCGTGCGCTCTGCGCGCTCCGCGCTAGGCTCCTCCGCGCTCACT 312

QY 991 CAGTACGGCGCGCGCTTCTGGGTACGCTGGCAACCGCG 1030
Db 313 CAGTACGGCGCGCGCTTCTGGGTACGCTGGCAACCGCG 352

Search completed: February 23, 2004, 23:36:10
Job time : 564.916 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 19:35:23 ; Search time 3666.04 Seconds
(without alignments)
11737.841 Million cell updates/sec

Title: US-09-936-456-1

Perfect score: 1441
Sequence: 1 aaagtaacggtacacagacag.....aaaaaaaaaaaaaaaa 1441

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	694.6	48.2	1594	11	AK018569 Mus muscu
2	657	45.6	1309	11	AK008816 Mus muscu
3	529.6	36.8	537	9	AI821606 nk08a11.x
4	515	35.7	515	9	AI826629 wk35e04.x

C	5	510	35.4	510	9	AI304327	AI304327 qo57h06.x
C	6	489.2	33.9	505	9	AI791844	AI791844 nk08a11.y
C	7	482.4	33.5	486	9	AI660560	AI660560 we68b08.x
C	8	466	32.3	1014	13	BY708714	BY708714 BY708714
C	9	446	31.0	446	9	AI984141	AI984141 wu21c02.x
C	10	445.4	30.9	466	9	AA573825	AA573825 nk08a11.s
C	11	434	30.1	434	9	AI991272	AI991272 wu41h04.x
C	12	414	28.7	414	9	AA593860	AA593860 nm19f03.s
C	13	409.4	28.4	411	9	AI983793	AI983793 wu20c09.x
C	14	405	28.1	405	9	AI274929	AI274929 q149c11.x
C	15	400.2	27.8	405	9	AI346155	AI346155 qp43f12.x
C	16	392	27.2	538	9	AI660493	AI660493 we67h02.x
C	17	388.4	27.0	538	13	BX517021	BX517021 BX517021
C	18	385	26.7	573	14	CB854121	CB854121 UI-CF-DU1
C	19	381	26.4	381	9	AI281211	AI281211 qk58e08.x
C	20	375.8	26.1	737	14	CB305399	CB305399 UI-CF-EN1
C	21	373.6	25.9	660	14	CF766127	CF766127 CES004870
C	22	372.8	25.9	646	13	BY720747	BY720747 BY720747
C	23	367.4	25.5	713	14	CB171450	CB171450 JID602700
C	24	356.4	24.7	1032	29	AY418362	AY418362 Homo sapi.
C	25	349.8	24.3	1026	29	AY418364	AY418364 Mus muscu
C	26	346	24.0	483	9	AI019726	AI019726 ua94c11.x
C	27	335	23.2	348	9	AW050605	AW050605 w219b11.x
C	28	333.4	23.1	690	12	BM977010	BM977010 UI-CF-EN1
C	29	333.4	23.1	896	12	BG968477	BG968477 602835393
C	30	324	22.5	687	14	CB850544	CB850544 UI-CF-EN1
C	31	322.4	22.4	324	12	BM987789	BM987789 UI-H-CO0-
C	32	308.8	21.4	871	12	BI526485	BI526485 602925070
C	33	304.8	21.2	919	12	BI412602	BI412602 602990526
C	34	296	20.5	296	9	AI821178	AI821178 ne17a03.y
C	35	295.2	20.5	999	12	BI410133	BI410133 602964073
C	36	290.6	20.2	315	9	AA618335	AA618335 nq15g11.s
C	37	289.4	20.1	895	12	BI413312	BI413312 602986494
C	38	275	19.2	277	9	AI732165	AI732165 ne17a03.x
C	39	274.4	19.1	608	14	CF108325	CF108325 Shultzomi
C	40	274.4	19.0	654	10	BE914798	BE914798 60167822
C	41	263.2	18.3	766	29	AY418363	AY418363 Pan trogl
C	42	260	18.0	260	9	AW009962	AW009962 ws89h06.x
C	43	257.2	17.8	675	12	BI154672	BI154672 602904653
C	44	252	17.5	252	9	AI262416	AI262416 qk38e04.x
C	45	251.8	17.5	1040	12	BI554670	BI554670 603236624

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
AK018569	LOCUS	AK018569	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	1	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	DEFINITION	Mus musculus adult male colon cDNA, RIKEN full-length enriched library, clone:9030623N16 product:hypothetical protein, full insert sequence.	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	2	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	ACCESSION	AK018569	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	3	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	VERSION	AK018569.1	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	4	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	KEYWORDS	HTC; CAP trapper.	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	5	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	SOURCE	Mus musculus (house mouse)	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	6	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	ORGANISM	Mus musculus (house mouse)	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	7	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	REFERENCE	1	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	8	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	AUTHORS	Carninci, P. and Hayashizaki, Y.	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	9	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	TITLE	High-efficiency full-length cDNA cloning	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	10	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	JOURNAL	JOURNAL	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	11	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	MEDLINE	MEDLINE	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	12	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	PUBMED	PUBMED	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	13	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	REFERENCE	2	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	14	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	AUTHORS	Carninci, P. and Hayashizaki, Y.	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	15	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	TITLE	High-efficiency full-length cDNA cloning	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	16	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	JOURNAL	JOURNAL	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	17	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	MEDLINE	MEDLINE	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	18	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning
AK018569	PUBMED	PUBMED	AK018569	AK018569.1	GI:12858338	HTC; CAP trapper.	Mus musculus (house mouse)	19	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning

REFERENCE
AUTHORS
3
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuina,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
20530913
11076861

REFERENCE
AUTHORS
4
The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

TITLE
JOURNAL
REFERENCE
AUTHORS
5
The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
REFERENCE
AUTHORS
6
(bases 1 to 1594)

Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
Hirooka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot = 10.0 and subtraction to
Rot = 185.2. Second strand cDNA was prepared with the primer
adapter of sequence [5'
GAGAGAGATCTCGAGTAAATTAATATCCCCCCCCCCC 3']. cDNA was cleaved
with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
bulk excision from lambda Fluc I. Cloning sites, 5' end: SalI, 3'
end: BamHI. Host: DH10B.
Location/Qualifiers
1..1594
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM,DB:9030623N16"
/db_xref="MGI:1894813"
/db_xref="taxon:10090"
/clone="9030623N16"
/sex="male"
/tissue_type="colon"

FEATURES
source

CDS
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
144..1106
/note="unnamed protein product; hypothetical protein
(evidence: Procest,decoder,NCBI CDS
Predictor,Longest-ORF)
putative"
/codon_start=1
/protein_id="BAB31281.1"
/db_xref="GI:12858339"
translation="MTAMDGVLPYQPRHAASPSVPLLVIVFLSLAASFILPG
IRGSRWFVLVRLVLSLFIGAEIVAVHPSGDMFVGRVWNTSYKAFSPSRVQVHGLH
VGLAGVNTLRGTPROLNETIDYNERFTRLNEDYTKYVHLEKLPDPLYAEK
FTPSPPCGLYHQYHLAAGHYAAATLVAFCEFWITANALLSMPADLYGLALLTGAFLL
EGVFAFASISSVPLGFRIGSAVLPYGAASFMLTATGILSLGGAVVILHYTRPS
ALRSFLDISVKDCSNQAKNSPLTLNPNQHEQLKSPDLNITLL"

ORIGIN
Query Match 48.2%; Score 694.6; DB 11; Length 1594;
Best Local Similarity 75.7%; Pred. No. 7.3e-85;
Matches 918; Conservative 0; Mismatches 279; Indels 16; Gaps 4;
222 CTGCCCGCTGCGTGCAGCACTGGCGCGGCGTGCAGCANGACCTGTGACGGCGTACT 281
Db CTGCTGCGCTGCGAAGCTTCTGCACTTCTGCAGANGACTGCGTGGACGGGGTCT 166
QY 282 GCCTTTTACCCCGAGCGCGCGATGCGGAGCTTCAAGCTTCACTGCTCATCGTTAT 341
Db 167 ACCCTTTACCCCGAGCGCGCGATGCGGAGCTTCAAGCTTCACTGCTCATCGTTAT 226
QY 342 TCTAGTGTGTTGGCTCTAGCAGAGAGCTTCTGCTCATCTTGGCGGATCCGTGCCA 401
Db 227 CCGTGTGTTCTGTGTTGCTTGGCTGTAGCTTCTGCTCATCTTGGCGGATCCGTGCCA 286
QY 402 CTCGCGGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 461
Db 287 CTCGCGGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 346
QY 462 GCGTGTGCACTTCACTGAGAGAGAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 521
Db 347 GCGTGTGCACTTCACTGAGAGAGAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 406
QY 522 AGCCTTACGCGAGCGCGCGCTTACAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 581
Db 407 AGCCTTACGCGAGCGCGCGCTTACAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 466
QY 582 TAATATTACACTACAGGAG 641
Db 467 TAACATTACACTCCAG 526
QY 642 GCAGTTACCTGCGCTTGAAGAAGAAATTAGCGCGCGAGTACGGAAGCACTGGAGAA 701
Db 527 GCGTTTCACTTGGCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 586
QY 702 GGGGTCGCGGAGCGAGTGTCTTACCTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 761
Db 587 GGGACTGCGGAGCGAGTGTCTTACCTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 646
QY 762 CCTGTACACAGTACAGCACTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 821
Db 647 GGTGTACACCAATATACAGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 706
QY 822 CTGCTTGTGCTCTCTTCAACAGTGTGCTTCAACGCGCGCGCGCTTCAACGAGAGGCT 881
Db 707 CTGCTTGTGATCATGCAATGCGCTGTGCTTCAATGCGCGCGCGCTTCAACGAGAGGCT 766
QY 882 GGCACGTGTGACACAG 941
Db 767 GGGCTTGTGATCAGACAG 826
QY 942 TAGCGTGGCTCTGCGCGCTTAGGCTTCCGCGCTACCACTACAGTACAGGCGC 1001
Db 827 CAGCGTGGCTCTGCGCGCTTCCGCGCTGGGCTCGCGCTTCCCTACAGGCGCTTACAGGCGC 886

QY	1002	CGCCTTCTGGGTCACGCTGGCAACCGGCGCTCTGTGCTCTTCTCTCGGAGGGGCGGTGT	1061
Db	887	CTCCTTTTGGCTCACGCTGGCCACCGGCATCTGAGCCTCTCTCTCGGAGGGGCGGTGT	946
QY	1062	GAGTCTCCAGTATGTTGGGCCACGCGCTCTTGCCACCCTTCTGGACCAAAAGCCCAAGGA	1121
Db	947	GATTCTCCACTATACTCGGCCACGCGCCTTGCTCTCTTTCTGATCTAAGTGTCAAAGA	1006
QY	1122	CT---GCAGCCAGAGAGAGGGGGCTCACCTCTTATCTCGGCGACCCACTGCAACAAGCA	1178
Db	1007	CTGTAGCAACCAAGGCTAAAGAAACTACCTCTCAACCTCAACAACCCGCAACGAACA	1066
QY	1179	GCGCGCTCTCCGAGACTTAAATGTATCACCACTAACCTGTGAGGGGACCAATCTGGA	1238
Db	1067	GTTGAAGAGTCCAGACTTAAA---TATTACACTCTCCTGTGAAATAACTGACTCCGGA	1123
QY	1239	CTCCTTCCCCCGCTTGGGACATCGCAGCGGGGAAGCAGTGCCTCCGACGCTG-GGCCA	1297
Db	1124	TTTCTACCCCTCTTTGGGACCCCATAGACCTGAAGCTGTGTTAAGCGCGTGCAGGA	1183
QY	1298	GGAGAGCTCCAGGAAGGGCAGCTGAAGCGCTGCTGGCGCGAGCGCTCGACATCCGACGGCA	1357
Db	1184	GCATAGCTTAGTCTGGAGAACTGTGTCCCTGCGGGACGACAAAGGGGGCACCCCATGTG	1243
QY	1358	CC-----AGGGAAGTCTCCTGGGGCGATCTGTAATPAAACCTTTTTCCTTTGT	1408
Db	1244	CTTTGCTCTAGAGAAATGATTTCTCAGGAGAACTGTAATPAAACCTTTTGTTCCTTT	1303
QY	1409	TTTTTAAAAAAA 1421	
Db	1304	TTTTTCTCAAAA 1316	

[illegible]

PUBMED	11076861
REFERENCE	4
AUTHORS	THE RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5
AUTHORS	THE FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
REFERENCE	6 (bases 1 to 1309)
AUTHORS	Adachi,T., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,

TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of

COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for

further details.
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues. First strand cDNA was primed with a primer
 [5' GAGGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse transcriptase
 and subsequently enriched for full-length by cap-trapper. Second
 strand cDNA was prepared with the primer adapter of sequence [5'
 GAGGAGAGAGATTCGAGTTATTAATTAATCCCCCCCC 3']. cDNA was cleaved
 with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
 Host: SOLR.

Location/Qualifiers

Source

CDS

```

1. 1309
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:2210403N03"
/db_xref="MGI:1893739"
/db_xref="taxon:10090"
/clone="2210403N03"
/sex="male"
/tissue_type="stomach"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
142. 1104
/note="unnamed protein product; hypothetical protein
(evidence: ProCrest, decoder, NCBI CDS
Predictor, Longest-ORF)
putative"
/codon_start=1
/protein_id="BAB5910.1"
/db_xref="GI:12843239"
/translation="MTAWDGVLPFYQPRHAASFVPLILVLPISLAASFILPFG
IRGSRWFWLVRVLLSLFTGABIVAVHFGSDWFVEEWTNTSYKAFSPRVQVHGLH
VGLGPGNLTIRGTPEQQLNETIDYNERFTWLNEDYREYVHAFEGLPDPVLYLAEK
FTSPSPGVTILRGVQVHLPQHYAAATLVGILLDHROCALHARPTLRGLGAHRCGRAR
LRCRFRPLDQFQALPLPPWMLRRLHALLRLLLTTLATGILSLLLGGAIVILHYTRPS

```

ORIGIN

ALRSLDLVSKDQSNQAKNSPLTLNPNQHEQLKSPDLNITTL"

Query Match 45.6%; Score 657; DB 11; Length 1309;
 Best Local Similarity 74.5%; Pred. No. 9.9e-80;
 Matches 898; Conservative 0; Mismatches 290; Indels 17; Gaps 5;

```

QY 222 CTGCCCCCTGCGTGCAGCACTCGGCGCGCGTGCAGCATGACCTGTGGAACGCGCTACT 281
Db 105 CTCGCTGCCTGCGAACAGTTCTGCAACTTCTTGACAGATGACTGCGTGGAACGCGGTGCT 164
QY 282 GCGTTTTCACCCCGAGCCCGGCGATGCGGAGCTTCAGCGTTCCACTGCTCATCTGTTAT 341
Db 165 ACCCTTTTACCCCGAGCCCGGAGGATGCGGAGCTTCAGCGTTACCGCTGCTCATCTGTTAT 224
QY 342 TCTAGTGTTTTGGCTCTAGCAGCAAGCTTCTGCTCATCTTGCCTGGGATCCGTGGCA 401
Db 225 CTTGTTGTTCTTGTCTTTGGCTGTAGCTTCTGTTCACTTGTGCTGGAATCCGTGGCA 284
QY 402 CTGCGCGTGGTTTGGTGTGAGAGTTCTTCTCATGTTCTCATAGGCGGAGAAATTGT 461
Db 285 CTCGCGCTGTTCTGTTGGTGTGAGAGTCTTCTTACCTGTTCTATAGGTGCGGAATTGT 344
QY 462 GCGTGTGCACTTCACTGCAGAAATGTTGTTGGTACAGTGAACACCAACATCTCTACAA 521
Db 345 GCGTGTGCACTTCACTGCAGAAATGTTGTTGGTGAAGATGTGGAACCAACATCTCTACAA 404
QY 522 AGCCTTCAGCGCAGCGCGCGTTACAGCCCGTGTGCTGCTGTTGGCGCTGAGGGCAT 581
Db 405 AGCCTTCAGTCCATCCCGTGTTCAGATCCATGCTGCTGCAAGTGGCGCGCGT 464
QY 582 TAATATTAACACTACAGGGAGCCCGAGTGCATGACTGAACGAGACCATTTGACTACACGA 641
Db 465 TAACATTAACACTCCGAGGAACACCGAGCAGCTGAACGAGACCATTTGACTACAAATGA 524
QY 642 GCAGTTACCTGCGCTGTGAAGAATAATTACCGCGCGGAGTACCGGAACGCACTGAGAA 701
Db 525 GCGTTTCACTTGGCGCTGTGAACGAGAACTACACCAAGAGTATGCCATGCTTTGAGAA 584
QY 702 GGGGCTGCGGAGACCGAGTGTCTACTTGGCGGAGAGTTCAACCCAGTAGCCCTTGCGG 761
Db 585 GGGGCTGCGGAGACCGAGTGTCTACTTGGCGGAGAGTTCAACCCAGTAGCCCTTGCGG 644
QY 762 CCTGTACCAACAGTACCACTGCGGAGACACTAGCGCTCGGCAAGCTATGGG-TGGCGT 820
Db 645 GGTGTACCAACATATCACTTGCAGGCTCATATGCGGAGAACACTGTGGGTTGCGAT 704
QY 821 TCTGCTTCTGGCTCTCTTCCACAGTGTCTCTCCAGCGCGCGCGCTCTACGAGAGCC 880
Db 705 TCTGCTTCTGGATCATGCGCAATGCGCTGTCTCCATGCGCGCGCGCTCTACGAGAGCT 764
QY 881 TGGCACTGTGACCAACCGAGCGCTTGGCGCTCTTGGGGTCTTGGCCTTGCCTCATCT 940
Db 765 TGGTTTGTCTACCAACCGAGTGTCTTGGCGCTCTTGGGTGTCTTGGCCTTGCCTCATCT 824
QY 941 CTAGCGTGCCTCTGCGCGCTCGCGCTAGGCTCTCTCGCGCTCAACCACTAGTAGCGCG 1000
Db 825 CCAGCGTGCCTCTGCGCACTCCCGCTTGGCTCGCGCTCTTCAACCGCTTACTACGCGG 884
QY 1001 CCGCCTTCTGGGTACCGTGTGCAACCGCGCTGTGCTGCTCTTCTCGAGGGCGCGTGG 1060
Db 885 CCTCCTT-TTGCTACCGTGTGCGCAACCGCGCATCTAGCGCTCTCTCTCGAGGGCGGTGG 943
QY 1061 TGAGTCTTCAGTATGTTGGCCCGAGCGCTCTTGCACCCCTTCTGGAACCAAGCGCAAGG 1120
Db 944 TGATTCTTCACTATACTGCGCCCGAGCGCTTCTGCTCTTCTTGTGATCTAAGTGTCAAAG 1003
QY 1121 ACT--GAGAGCCAGAGAGAGGGGCTCACTCTTATCTCTGGCGAACCCACTGCAACAAGC 1177
Db 1004 ACTGTAGCAACAGGCTAAAGAACTCACTCTCAACCTCAACAACCCGCAACAGAAC 1063
QY 1178 AGGCGGCTTCCAGACTTAAATGTATCAACCACTTGTGAGGGGAGCCCAATCTGG 1237

```

```

Db 1064 AGTTGAAGAGTCCAGACTTAA--TATTACCACTCTCTGTGAAATACTGACTCCGG 1120
QY 1238 ACTCCTTCCCGCCCTTGGGACATGCGAGCGCCGGAGACAGTCCCGCCAGGCTGCGCCA 1297
Db 1121 ATTCTTACCCCTCTTTGGGACCCCATAGACCTGAAGACTGTGTTAAGAGCGCGTGCAGG 1180
QY 1298 GGAGAGCTCCAGAGAGGGCAGCTGAGCGCTGTGCGCGGAGGCGCTCGACATCCGAGGCA 1357
Db 1181 AGCATAGCTTAGTCTGGAGAACTGTCTTCTGCGCGGAGAGCAAGGGGACCCCATGTG 1240
QY 1358 CC-----AGGAAAGTCTCCTGGGCGCATCTGTAAATAACCTTTTCTTTTGT 1408
Db 1241 CCTTGTCTTAGAGATGATTTCTCAGAGAACTGTAAATGAACCTTTTGTCTTTT 1300
QY 1409 TTTT 1413
Db 1301 TTTT 1305

```

RESULT 3
 A1821606/c
 LOCUS
 DEFINITION
 A1821606 537 bp mRNA linear EST 13-DEC-1999
 similar to contains_TARI.t3 TARI MER22 repetitive element ;, mRNA
 sequence.

ACCESSION A1821606 GI:5440685
 VERSION A1821606.1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS NCI-CGAP
 TITLE Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Other ESTs: nk08a11.y5

CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Stratagene, Inc., David B. Krizman,
 Ph.D.

CDNA Library Arraying: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 www-bio.lnl.gov/bbrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
 Original clone citation: National Cancer Institute, Cancer Genome
 Anatomy Project (CGAP), Tumor Gene Index
 This read has been verified (found to hit its original self in the
 correct orientation)
 Insert Length: 741 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 459.

FEATURES

source

1. 537
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1012892"
 /issue_type="tumor"
 /lab_host="SOLR (kanamycin resistant)"
 /clone_lib="NCI CGAP Co2"
 /note="Organ: colon; Vector: Bluescript SK-; Site 1:
 EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. Bulk colon villous adenoma. 5' adaptor sequence:
 5' GAATTGGGACGAG 3' 3' adaptor sequence: 5'
 CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

ORIGIN

Query Match 36.8%; Score 529.6; DB 9; Length 537;
 Best Local Similarity 99.3%; Pred. No. 2.8e-62;
 Matches 532; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 878 GCCTGGCACTGCTGACCAACCGAGCCTTCGCGCTCTTCGGGGCTTCGCTTGCCCTTCA 937
 |||||
 DB 537 GCCTGGCACTGCTGACCAACCGAGCCTTCGCGCTCTTCGGGGCTTCGCTTGCCCTCA 478
 |||||

QY 938 TCTCTAGCGTGCCGCTCTGCTCCCGCTAGGCTCCTCCGGCTCACCACCTCAGTACG 997
 |||||
 DB 477 TCTCTAGCGTGCCGCTCTGCTCCCGCTAGGCTCCTCCGGCTCACCACCTCAGTACG 418
 |||||

QY 998 GCGCCGCTTCTGGGTCAAGCTGGCAACCGCGCTGCTGCTCTTCTCTCGAGGGGCGG 1057
 |||||
 DB 417 GCGCCGCTTCTGGGTCAAGCTGGCAACCGCGCTGCTGCTCTTCTCTCGAGGGGCGG 358
 |||||

QY 1058 TGGTGAATCTCCAGTATGTTGGGCCCAAGCGCTCTTCGCAACCTTCTTGACCAAGCGCCA 1117
 |||||
 DB 357 TGGTGAATCTCCAGTATGTTGGGCCCAAGCGCTCTTCGCAACCTTCTTGACCAAGCGCCA 298
 |||||

QY 1118 AGGACTGCAGCCAGGAGAGAGGGGGCTCACCCTTTATCTCTCGGCGACCCACTGCACAAGC 1177
 |||||
 DB 297 AGGACTGCAGCCAGGAGAGAGGGGGCTCACCCTTTATCTCTCGGCGACCCACTGCACAAGC 238
 |||||

QY 1178 AGCCCGCTCTCCAGACTTAAATGATATCACCACCTGTGAGGGGGACCAATCTGG 1237
 |||||
 DB 237 AGCCCGCTCTCCAGACTTAAATGATATCACCACCTGTGAGGGGGACCAATCTGG 178
 |||||

QY 1238 ACTCCTTCCCCCGCTTGGGACATCCGAGGCCGGAAGCAGTGCCTCCGAGCTGGGCCA 1297
 |||||
 DB 177 ACTCCTTCCCCCGCTTGGGACATCCGAGGCCGGAAGCAGTGCCTCCGAGCTGGGCCA 118
 |||||

QY 1298 GGAGAGCTCCAGGAAGGCACTGAGCGCTGCTGGCGGAGGCTCCGACATCCGAGGCA 1357
 |||||
 DB 117 GGAGAGCTCCAGGAAGGCACTGAGCGCTGCTGGCGGAGGCTCCGACATCCGAGGCA 58
 |||||

QY 1358 CCAGGGAAGTCTCTCGGGCGGATCTGTAATAAACCCTTTTCTTTTCTTTT 1413
 |||||
 DB 57 CCAGGGAAGTCTCTCGGGCGGATCTGTAATAAACCCTTTTCTTTTCTTTATGATTT 2
 |||||

RESULT 4
 AI826629 515 bp mRNA linear EST 21-DEC-1999
 LOCUS WK35e04.x1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:2417406 3',
 DEFINITION mRNA sequence.
 ACCESSION AI826629
 VERSION AI826629.1 GI:5447300
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 515)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 910 Std Error: 0.00
 Seg primer: -40UP from Gibco
 High quality sequence stop: 467.
 Location/Qualifiers

source 1. .515
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2417406"
 /sex="male"
 /tissue_type="normal prostate"
 /lab_host="DH10B"
 /clone_lib="NCI_CGAP_Pr22"
 /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
 with a modified polylinker; 1st strand cDNA was prepared
 from normal prostate bulk tissue, and was then primed with
 a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pT7T3 vector. Library is normalized, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 35.7%; Score 515; DB 9; Length 515;
 Best Local Similarity 100.0%; Pred. No. 2.7e-60;
 Matches 515; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 902 CCTTGGCGCTCTTGGGGCTTGGCCTTGCATCTCTAGCGTCCGCTCTGCCCCG 961
 |||||
 DB 515 CCTTGGCGCTCTTGGGGCTTGGCCTTGCATCTCTAGCGTCCGCTCTGCCCCG 456
 |||||

QY 962 TCCGCTAGGCTCTCCGGCTCACCACCTAGTACGGCGCCGCTTCTGGGTCAAGCTGG 1021
 |||||
 DB 455 TCCGCTAGGCTCTCCGGCTCACCACCTAGTACGGCGCCGCTTCTGGGTCAAGCTGG 396
 |||||

QY 1022 CAACCGGCGCTGTGCTCTTCTCTCGAGGGGGCGGTGAGTCTCCAGTATGTTGGC 1081
 |||||
 DB 395 CAACCGGCGCTGTGCTCTTCTCTCGAGGGGGCGGTGAGTCTCCAGTATGTTGGC 336
 |||||

QY 1082 CAGCGCTCTTGCACCCCTTCTGACCAAGCGCCAAAGACTGACGCGAGAGAGGGG 1141
 |||||
 DB 335 CAGCGCTCTTGCACCCCTTCTGACCAAGCGCCAAAGACTGACGCGAGAGAGGGG 276
 |||||

QY 1142 GCTCACCCTTATCTCTCGGAGCCACCTGACCAAGAGCGGCTCTCCAGACTTAAAT 1201
 |||||
 DB 275 GCTCACCCTTATCTCTCGGAGCCACCTGACCAAGAGCGGCTCTCCAGACTTAAAT 216
 |||||

QY 1202 GTATCACCACCTGATGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGGACATC 1261
 |||||
 DB 215 GTATCACCACCTGATGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGGACATC 156
 |||||

QY 1262 GCAGGCGGGAAGCAGTGCCTCCAGGCTGGGCGAGAGAGCTCCAGGAAGGCACTGA 1321
 |||||
 DB 155 GCAGGCGGGAAGCAGTGCCTCCAGGCTGGGCGAGAGAGCTCCAGGAAGGCACTGA 96
 |||||

QY 1322 GCGCTGCTGGCGAGGCTTCGACATCCGAGGACCAAGGAAAGTCTCTGGGGCAT 1381
 |||||
 DB 95 GCGCTGCTGGCGAGGCTTCGACATCCGAGGACCAAGGAAAGTCTCTGGGGCAT 36
 |||||

QY 1382 CTGTAATTAACCTTTTCTTTCTTTGTTTAA 1416
 |||||
 DB 35 CTGTAATTAACCTTTTCTTTCTTTGTTTAA 1
 |||||

RESULT 5
 AI304327 510 bp mRNA linear EST 01-FEB-1999
 LOCUS AI304327/c
 DEFINITION q057h06.x1 NCI_CGAP_C08 Homo sapiens cDNA clone IMAGE:1912667 3',
 similar to contains TARI.t3 TARI repetitive element ;, mRNA
 sequence.
 ACCESSION AI304327
 VERSION AI304327.1 GI:3988016
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 510)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www-bio.lnl.gov/bbrp/image/image.html
Insert length: 1487 Std Error: 0.00
Seq primer: -40up from Gibco
High quality sequence stop: 476.
Location/Qualifiers

FEATURES
source

1. 510
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1912667"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI CGAP CO8"
/note="Organ: colon; Vector: pT7AD-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 35.4%; Score 510; DB 9; Length 510;
Best Local Similarity 100.0%; Pred. No. 1.3e-59;
Matches 510; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 904 TTGCGCTCTTGGGGGTTCTTGGCTTGGCTTCCATCTAGCGTGGCGGCTGCGCCGCTC 963
Db 510 TTGCGCTCTTGGGGGTTCTTGGCTTGGCTTCCATCTAGCGTGGCGGCTGCGCCGCTC 451
QY 964 CGCCTAGGCTCTCGCGCTCAGCACTAGTACGGCGCGCTTCTGGGTACAGTGGCA 1023
Db 450 CGCCTAGGCTCTCGCGCTCAGCACTAGTACGGCGCGCTTCTGGGTACAGTGGCA 391
QY 1024 ACCGGCTCTGCTGCTCTTCTCTCGAGGGGGCGGTGTGAGTCTCCAGTATGTTGCGCC 1083
Db 390 ACCGGCTCTGCTGCTCTTCTCTCGAGGGGGCGGTGTGAGTCTCCAGTATGTTGCGCC 331
QY 1084 AGCGCTCTTGGACCCCTTCTGACCAAGCGCCAGAGACTGCAAGCAGAGAGAGAGGGGC 1143
Db 330 AGCGCTCTTGGACCCCTTCTGACCAAGCGCCAGAGACTGCAAGCAGAGAGAGAGGGGC 271
QY 1144 TCACCTCTTATCTTGGGCGACCACTGCACAGAGAGCGCTCTCCAGACTTAAATGT 1203
Db 270 TCACCTCTTATCTTGGGCGACCACTGCACAGAGAGCGCTCTCCAGACTTAAATGT 211
QY 1204 ATCACCACCTAAGCTGTGAGGGGGAGCCCAATCTGSACTCCTTCCCGCTTGGGACATGC 1263
Db 210 ATCACCACCTAAGCTGTGAGGGGGAGCCCAATCTGSACTCCTTCCCGCTTGGGACATGC 151
QY 1264 AGCGCGGGAAGAGTGCCTCCGAGGCTTGGGCGAGAGAGTCCAGGAAGGCACTGAGC 1323
Db 150 AGCGCGGGAAGAGTGCCTCCGAGGCTTGGGCGAGAGAGTCCAGGAAGGCACTGAGC 91
QY 1324 GCTGCTGGCGGAGGCTCGACATCCGAGGCAAGGGAAGTCTCTGGGGCGATCT 1383
Db 90 GCTGCTGGCGGAGGCTCGACATCCGAGGCAAGGGAAGTCTCTCTGGGGCGATCT 31

QY 1384 GTAAATPAACCTTTTCTTTCTTTGTTTTT 1413
Db 30 GTAAATPAACCTTTTCTTTCTTTGTTTTT 1

RESULT 6
AI791844 505 bp mRNA linear EST 13-DEC-1999
LOCUS nk08a11.y5 NCI-CGAP CO2 Homo sapiens cDNA clone IMAGE:1012892 5',
DEFINITION similar to contains Alu repetitive element, mRNA sequence.
ACCESSION AI791844
VERSION AI791844.1 GI:5339486
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 505)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Other ESTs: nk08a11.s1
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc., David B. Krizman,
Ph.D.

CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www-bio.lnl.gov/bbrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: see original entry for original citation
information
This 5' resequenced clone has no previous 5' data to verify this
new read against
Putative full length read
The vector to vector length is 671
Insert length: 741 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 418.
Location/Qualifiers

FEATURES
source

1. 505
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1012892"
/tissue_type="tumor"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="NCI CGAP CO2"
/note="Organ: colon; Vector: Bluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Bulk colon villous adenoma. 5' adaptor sequence:
5' GAATTCGACACGAG 3' 3' adaptor sequence: 5'
CTCAGTTTCTTTTCTTTTCTTTT 3' Average insert size: 1.1 kb."

ORIGIN

Query Match 33.9%; Score 489.2; DB 9; Length 505;
Best Local Similarity 99.2%; Pred. No. 8.3e-57;
Matches 502; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 736 AAGTTACACCGAGTAGCCCTTGGCGCTGTACACCAAGTACCACTGGCGGACACTAC 795
Db 1 AAGTTACACCGAGTAGCCCTTGGCGCTGTACACCAAGTACCACTGGCGGACACTAC 60
QY 796 GCCTCGGCACGCTATGGGTGGCGTCTGCTTCTGAGCTCCTCTCAACGTGCTCTCC 855
Db 61 GCCTCGGCACGCTATGGGTGGCGTCTGCTTCTGAGCTCCTCTCAACGTGCTCTCC 120

```

QY 856 ACGCGCGCGCGCTCTACGAGAGCGCTGGCACTGCTGACCACCGGAGCCTTCGGCTCTTC 915
Db 121 ACGCGCGCGCGCTCTACGAGAGCGCTGGCACTGCTGACCACCGGAGCCTTCGGCTCTTC 180
QY 916 GGGGTCTTCGCTTGGCTTCCATCTCTAGCGGTGCGGCTCTGCGGCTCCGCTAGGCTCC 975
Db 181 GGGGTCTTCGCTTGGCTTCCATCTCTAGCGGTGCGGCTCTGCGGCTCCGCTAGGCTCC 240
QY 976 TCCGCGCTCACCACTCACTAGACGCGCGCGCTTCTGGGTCAAGCTGGCAACCGGCGTCC 1035
Db 241 TCCGCGCTCACCACTCACTAGACGCGCGCGCTTCTGGGTCAAGCTGGCAACCGGCGTCC 300
QY 1036 TGCCTCTTCTTCGAGAGGCGCGCTGTGTAGTCTCCAGTATGTTCCGCGCCAGCGCTCTTC 1095
Db 301 TGCCTCTTCTTCGAGAGGCGCGCTGTGTAGTCTCCAGTATGTTCCGCGCCAGCGCTCTTC 360
QY 1096 ACCCTTCTGAGACCAAGCGCCAAAGAGTGCAGCCAGAGAGAGAGGCGGCTCACTCTTATC 1155
Db 361 ACCCTTCTGAGACCAAGCGCCAAAGAGTGCAGCCAGAGAGAGAGGCGGCTCACTCTTATC 420
QY 1156 CTGCGGCACTGCAAGAGCGCGCTCTCCAGACTTAAGTATACCACTAAC 1215
Db 421 CTGCGGCACTGCAAGAGCGCGCTCTCCAGACTT-AAATGATACCACTAAC 479
QY 1216 CTGTAGGGGAGCCCAATCTGACTC 1241
Db 480 CTGTAGGGGAGCCCAATCTGACTC 505

```

RESULT 7

```

Al660560/c 486 bp mRNA linear EST 18-DEC-1999
LOCUS we68b08.x1 Soares Dieckgraefe colon NHCD Homo sapiens cDNA clone
DEFINITION IMAGE:2346231.3, similar to contains TAR1.t3 TAR1 repetitive
element, mRNA sequence.

```

```

ACCESSION Al660560
VERSION Al660560.1 GI:4764130
KEYWORDS EST.
SOURCE Homo sapiens (human)

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE 1 (bases 1 to 486)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

```

```

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

```

```

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

```

```

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.

```

```

Insert Length: 1433 Std Error: 0.00
Seq primer: -40UP from Gibco

```

```

High quality sequence stop: 447.
Location/Qualifiers

```

FEATURES

```

source

```

```

1..486
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2346231"
/tissue_type="colonic mucosa from 3 patients with Crohn's
disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares_Dieckgraefe_colon_NHCD"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGCTTTTCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Tissue samples
provided by Dr. Brian Dieckgraefe (Washington University,

```

```

dieck@im.wustl.edu); colonic mucosa represents a range of
disease involvement from moderate to severe Crohn's
disease; samples include both perforating (fistulas) and
non-perforating samples. Library constructed by Bento
Soares and M. Fatima Bonaldo. "

```

ORIGIN

```

Query Match 33.5%; Score 482.4; DB 9; Length 486;
Best Local Similarity 99.4%; Pred. No. 7e-56;
Matches 483; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 943 ACGGTGCGCTCTGCGCGCTCCGCTAGGCTCTCCGCGCTCACCACTAGTACGGCGCC 1002
Db 486 ACGGTGCGCTCTGCGCGCTCCGCTAGGCTCTCCGCGCTCACCACTAGTACGGCGCC 427
QY 1003 GCGTTCTGGGTACGCTGGCAACCGGCGTCTGTGCTCTTCTCTGAGAGGGCGGTGTG 1062
Db 426 GCGTTNTGGGTACGCTGGCAACCGGCGTCTGTGCTCTTCTCTGAGAGGGCGGTGTG 367
QY 1063 AGTCTCCAGTATGTTGGGCGGCGGCTCTTTCGCAACCTTCTGACCAAGCGCCAGGAC 1122
Db 366 AGTCTCCAGTATGTTGGGCGGCGGCTCTTTCGCAACCTTCTGACCAAGCGCCAGGAC 307
QY 1123 TGCAGCCAGAGAGAGAGGCGGCTCACTCTTATCTCTGCGGCAACCACTGACAGAGGCC 1182
Db 306 TGCAGCCAGAGAGAGAGGCGGCTCACTCTTATCTCTGCGGCAACCACTGACAGAGGCC 247
QY 1183 GCTCTCCAGACTTAAATATATATACCACTAACCTGTGAGGGGAGCCATCTGACTCC 1242
Db 246 GCTCTCCAGACTTAAATATATATACCACTAACCTGTGAGGGGAGCCATCTGACTCC 187
QY 1243 TTCCCGCGCTTGGGACATCGCAGAGCGCGGAGAGCAGTGGCCCGCCAGGCTGGGCGCAGAGA 1302
Db 186 TTCCCGCGCTTGGGACATCGCAGAGCGCGGAGAGCAGTGGCCCGCCAGGCTGGGCGCAGAGA 127
QY 1303 GCTCCAGAGAGGCACTAGAGCGCTGCTGGCGCGGAGCGCTCGGACATCGCAGAGCAGG 1362
Db 126 GCTCCAGAGAGGCACTAGAGCGCTGCTGGCGCGGAGCGCTCGGACATCGCAGAGCAGG 67
QY 1363 GAAAGTCTCTGGGCGGATCTGTAATAAACCCTTTTCTTTTGTAAAAA 1422
Db 66 GAAAGTCTCTGGGCGGATCTGTAATAAACCCTTTTCTTTTGTAAAAA 7
QY 1423 AAAAAA 1428
Db 6 AAAAAA 1

```

RESULT 8

```

BY708714 1014 bp mRNA linear EST 16-DEC-2002
LOCUS BY708714 RIKEN full-length enriched, adult male stomach Mus
DEFINITION musculus cDNA clone 2210403N03 5', mRNA sequence.

```

```

ACCESSION BY708714
VERSION BY708714.1 GI:27119906
KEYWORDS EST.
SOURCE Mus musculus (house mouse)

```

```

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE 1 (bases 1 to 1014)

```

```

AUTHORS

```

```

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaoka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schobach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Brdiczka, V.,
Chochia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Fraser, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, B.D., Kanai, A.,
Kawaji, H., Kawasaki, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,

```


Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Sempke,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

Contact: Yoshinide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watabiki,A., Muramatsu,M. and Hayashizaki,Y.
Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

Location/Qualifiers
1. 1014

FEATURES
source

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="2210403N03"

/sex="male"

/tissue_type="stomach"

/dev_stage="adult"

/lab_host="SOLR"

/clone_1b="RIKEN full-length enriched, adult male stomach"

/note="Site 1: XhoI; Site 2: SctI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was

primed with a primer [5'
GAGAGAGAGCGCGCGCAACTGAGTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGATTCTCGATTATTAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SctI."

ORIGIN

Query Match 32.3%; Score 466; DB 13; Length 1014;
Best Local Similarity 80.6%; Pred. No. 7,2e-54;
Matches 569; Conservative 0; Mismatches 135; Indels 2; Gaps 2;

QY	222	CTGCCCGCTGCGTGAGCACTGCGCGCGCGTGAGCATGACCTGTGGAACGGCGTACT	281
DB	105	CTGCTGCGCTGCGAAGTCTGCAACTCTTGACAGATGATGCGTGGGACGGGGTGCT	164
QY	282	GCCTTTTACCCCGAGCCCGGCGATGCGGAGCTTCAGCGTTCCACTGCTCATCGTTAT	341
DB	165	ACCCCTTTACCCCGAGCCCGGCGATGCGGAGCTTCAGCGTTCCACTGCTCATCGTTAT	224
QY	342	TCTAGTGTGTTGGCTCTAGCAGCAAGCTTCTGCTCATCTTGGCGGGATCCGTGGCCA	401
DB	225	CCTGATGTTCTGTTCTTTGGCTGCTAGCTTCTGCTCATCTTGGCGGGATCCGTGGCCA	284
QY	402	CTGCGCGCTGTTTGGTTGGTGAGAGTCTTCTGAGTCTGCTCATAGGCGGAGAAATTGT	461
DB	285	CTGCGCGCTGTTTGGTTGGTGAGAGTCTTCTGAGTCTGCTCATAGGCGGAGAAATTGT	344
QY	462	GGCTGTGCACTTCAGTGCAAGATGTTGTTGGTTGAGTGAACACCAACATCTTACAA	521
DB	345	GGCTGTGCACTTCAGTGCAAGATGTTGTTGGTTGAGTGAACACCAACATCTTACAA	404
QY	522	AGCCTTACGCGGAGCGCGCTTACAGCCCGTGTGCTGCTGCTGGGCTTGAAGGGCAT	581
DB	405	AGCCTTACGCTCATCCCGGTGTTCAAGTCCATGTCGTCAGTGGGCTGGCGCGCGCT	464
QY	582	TAATATTACACTCACAGGAGACCCGAGTGCATGAGTGAACGAGACCATGATACACGA	641
DB	465	TAACATTACACTCCGAGAACACCCGAGAGAGTGAACGAGACCATGATACACATGA	524
QY	642	GCAGTTCACCTGGCGTGAAGAGATTTAGCGCGGAGTAGCGGAGCACTGGAGAA	701
DB	525	GCCTTTCACCTGGCGTGAAGAGATTTAGCGCGGAGTAGCGGAGCACTGGAGAA	584
QY	702	GGGGCTGCGGAGCCAGTGTCTACTGCGGAGAGATTACACCGAGTAGCCCTTGGCG	761
DB	585	GGGGCTGCGGAGCCAGTGTCTACTGCGGAGAGATTACACCGAGTAGCCCTTGGCG	644
QY	762	CCTGTACACAGTACACACTGCGGAGACACTAGCCTGCGGACGCT-ATGGGTGGCGT	820
DB	645	GCTGTACACAGTATATACCTGCGGAGTCAATGCGGAGGACACTGTTGAGGTGCAT	704
QY	821	TCTGCTTCTGGCTCTCTCCAAAGTGTGCTGTCCACGCGCGCGCTCTACGAGAGCC	880
DB	705	TCTGCTTCTGGATCATGACCATATGCGGTGCTCTCCATGCGGCACTCTACGAGAGCC	764
QY	881	TGGCACTGTGACCAACGAGGCTTGGCGCTCTTGGGGGTCTTGGC	926
DB	765	TGGCTTGTCTACCAACC-GTGCCTTGGCGCTCTTGGGTGTCTTCC	809

RESULT 9

AI984141/c

LOCUS

AI984141 446 bp mRNA linear EST 27-OCT-1999

DEFINITION

wu21c02.x1 Soares Dieckgraefe colon_NHCD Homo sapiens cDNA clone

IMAGE:2520674 3', mRNA sequence.

ACCESSION

AI984141

VERSION

AI984141.1

KEYWORDS

EST.

SOURCE

ORGANISM

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 446)

AUTHORS

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco.

FEATURES

source

1..446
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2520674"
/tissue_type="colonic mucosa from 3 patients with Crohn's disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares Dieckgraebe colon NHCD"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACCAATCTGAAGTGGAGAGCGGCCGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraebe (Washington University, dieck@wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 31.0%; Score 446; DB 9; Length 446;
Best Local Similarity 100.0%; Pred. No. 6.1e-51;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 971 GCTCCTCCGCGCTCACCACCTCAGTACGGCGCGCCTTCTGGTCAAGCTGGCAACCGGCG
DB 446 GCTCCTCCGCGCTCACCACCTCAGTACGGCGCGCCTTCTGGTCAAGCTGGCAACCGGCG 387
QY 1031 TCCTGTGCGCTTCTCCTCGGAGGGCGGTGTGAGTCTCCAGTATGTTCCGCCCAAGCGCTC 1090
DB 386 TCCTGTGCGCTTCTCCTCGGAGGGCGGTGTGAGTCTCCAGTATGTTCCGCCCAAGCGCTC 327
QY 1091 TTGCAACCTTCTGACCAAGCGCCAGAGCTGCAGCCAGAGAGAGAGGGGCTCACCTC 1150
DB 326 TTGCAACCTTCTGACCAAGCGCCAGAGCTGCAGCCAGAGAGAGAGGGGCTCACCTC 267
QY 1151 TTATCCTCGGCGGACCACTGCACAGCAGCGCGCTTCCAGACTTAATATGATACCA 1210
DB 266 TTATCCTCGGCGGACCACTGCACAGCAGCGCGCTTCCAGACTTAATATGATACCA 207
QY 1211 CTAACCTGTGAGGGGAGCCCAATCTGACTCCTTCCCGCCTTGGAGACATCGAGCGCGG 1270
DB 206 CTAACCTGTGAGGGGAGCCCAATCTGACTCCTTCCCGCCTTGGAGACATCGAGCGCGG 147
QY 1271 GAAGCAGTGCCTCGGCGGCTGGGCGGAGAGAGCTCCAGAAAGGCACTGAGCGCTGCTG 1330
DB 146 GAAGCAGTGCCTCGGCGGCTGGGCGGAGAGAGCTCCAGAAAGGCACTGAGCGCTGCTG 87
QY 1331 GCGCGAGGCTCGGACATCCGAGGCAAGGAAAGTCTCTGGGGGATCTGTAAATA 1390
DB 86 GCGCGAGGCTCGGACATCCGAGGCAAGGAAAGTCTCTGGGGGATCTGTAAATA 27
QY 1391 AACCTTTTCTTTTGTGTTTAA 1416
DB 26 AACCTTTTCTTTTGTGTTTAA 1

RESULT 10

AA573825/c 466 bp mRNA linear EST 12-SEP-1997
LOCUS nX08a11.s1 NCI_CGAP_Co2 Homo sapiens cDNA clone IMAGE:1012892 3', mRNA sequence.

DEFINITION

AA573825
AA573825.1 GI:2348340

ACCESSION

AA573825
AA573825.1 GI:2348340

VERSION

EST.

KEYWORDS

Homo sapiens (human)

SOURCE

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

AUTHORS

Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmer-Buck, M.D., Ph.D.
cDNA Library Preparation: Stratagene, Inc., David B. Krizman, Ph.D.

JOURNAL

Unpublished (1997)

COMMENT

cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 741 Std Error: 0.00
Seq primer: -40m13 fwd. BT from Amersham
High quality sequence stop: 456.

FEATURES

source

1..466
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1012892"
/tissue_type="tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: colon; Vector: Bluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Bulk colon villous adenoma. 5' adaptor sequence: 5' GAATTCGACGAG 3' 3' adaptor sequence: 5' CTGAGTCTTTTCTTTTCTTTT 3' Average insert size: 1.1 kb."

ORIGIN

Query Match 30.9%; Score 445.4; DB 9; Length 466;
Best Local Similarity 98.5%; Pred. No. 7.1e-51;
Matches 460; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 947 TGCCGCTCTGCGCGCTCCGCGCTAGGCTTCTCCGCGCTCACCACCTAGTACGGCGCGGCT 1006
DB 466 TGCCGCTCTGCGCGCTCCGCGCTAGGCTTCTCCGCGCTCACCACCTAGTACGGCGCGGCT 407
QY 1007 TCTGGGTACGCTGGCAACCGGCGTCTGTGCTCTTCTCTCGAGGGCGGTGTGAGTC 1066
DB 406 TCTGGGTACGCTGGCAACCGGCGTCTGTGCTCTTCTCTCGAGGGCGGTGTGAGTC 347
QY 1067 TCCAGTATGTTCCGCCAGCGCTCTTGCACACCTTCTGACCAAGCGCCAGAGACTGCA 1126
DB 346 TCCAGTATGTTCCGCCAGCGCTCTTGCACACCTTCTGACCAAGCGCCAGAGACTGCA 287
QY 1127 GCCAGAGAGAGGGGCTCACTCTTATCTCTGGCGACCCACTGCACAAGAGCGGCTC 1186
DB 286 GCCAGAGAGAGGGGCTCACTCTTATCTCTGGCGACCCACTGCACAAGAGCGGCTC 227
QY 1187 TCCAGACTTAATATGTATCACCACCTAAGTGTAGGGGAGACCAATCTGACTCCTCC 1246
DB 226 TCCAGACTTAATATGTATCACCACCTAAGTGTAGGGGAGACCAATCTGACTCCTCC 167

QY	1247	CCGCGCTTGGGACATCCGACAGGCCGGGAAAGCAGTGTCCCGCAGGCGCTGGGGCCAGAGAGCTC	1306
Db	166	CCGCGCTTGGGACATCCGACAGGCCGGGAAAGCAGTGTCCCGCAGGCGCTGGGGCCAGAGAGCTC	107
QY	1307	CAGGAAAGGGCAGCTGAGCGCTGTGGCGCGAGGCGCTCGACATCCGACAGGCAACGAGGAAA	1366
Db	106	CAGGAAAGGGCAGCTGAGCGCTG-TGGCGGAGGCGCTCGACATCCGACAGGCAACGAGGAAA	48
QY	1367	GTCCTCTGGGGCGATCTGTAAATAAACCTTTTTTCTTTTGTTTTT	1413
Db	47	GTCCTCTGGGGCGATCTGTAAATAAACCTTTTTTCTTTTGTTTTT	1

RESULT 11						
AI991272/c						
LOCUS	AI991272	434 bp	mRNA	linear	EST 09-MAR-2000	
DEFINITION	wu41h04.x1 Soares Dieckgrafe_colon_NHCD Homo sapiens cDNA clone					
IMAGE:2522647 3'						mRNA sequence.

ACCESSION A1991272
VERSION A1991272.1 GI:5836177
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 434)	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 558 Std Error: 0.00
Seq primer: -40UP from Gibco.

FEATURES	Location/Qualifiers
source	1. .434

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2522647"
/tissue_type="colonic mucosa from 3 patients with Crohn's
disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares_Dieckgraeffe_colon_NHCD"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCATCTGGAAGTGGGAGCGCGCCGCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Tissue samples
provided by Dr. Brian Dieckgraeffe (Washington University,
dieck@im.wustl.edu); colonic mucosa represents a range of
disease involvement from moderate to severe Crohn's
disease; samples include both perforating (fistulas) and
non-perforating samples. Library constructed by Bento
Soares and M. Fatima Ronaldo."

```

ORIGIN

```

Query Match          30.1%; Score 434; DB 9; Length 434;
Best Local Similarity 100.0%; Pred. No. 2.6e-49;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0

```

QY	981	GCTACCACTCAGTACGGCGCCGCTTCTTGAGTCACGCTGGCAA	CGGCGTCTGTGCCCT	1040
Db	434	GCTACCACTCAGTACGGCGCCGCTTCTGGGTACGCTGGCAA	CGGCGTCTGTGCCCT	375

[illegible]

QY	1101	TCTGGACCAAAAGCCGCCAAGGACTGCAGCCAGGAGAGAGGGGGCTCA CCTCTTATCTCGG	1160
Db	314	TCTGGACCAAAAGCCGCCAAGGACTGCAGCCAGGAGAGAGGGGGCTCA CCTCTTATCTCGG	255
QY	1161	CGACCCCACTGCACAAGCAGGCGGCTCTCCAGACTTAAATGTATCA CCACTAACCTGTG	1220
Db	254	CGACCCCACTGCACAAGCAGGCGGCTCTCCAGACTTAAATGTATCA CCACTAACCTGTG	195
QY	1221	AGGGGGAGCCCAATCTGGACTCTCTCCCGGCTTGGGACATCGCAGGCGGGAGCACTGC	1280
Db	194	AGGGGGAGCCCAATCTGGACTCTCTCCCGGCTTGGGACATCGCAGGCGGGAGCACTGC	135
QY	1281	CCGCCAGGCGCTGGGGCCAGGAGAGCTCCAGGAAAGGGCACTGAGCGCTGTGGCGCGAGGCC	1340
Db	134	CCGCCAGGCGCTGGGGCCAGGAGAGCTCCAGGAAAGGGCACTGAGCGCTGTGGCGCGAGGCC	75
QY	1341	TCCGACATCCGCAGGACCGACCAAGGAAAGTCTCTCTGGGGCGATCTGTAAATAAACCTTTTTT	1400
Db	74	TCCGACATCCGCAGGACCGACCAAGGAAAGTCTCTCTGGGGCGATCTGTAAATAAACCTTTTTT	15
QY	1401	TCCTTTTGTTTTTT 1414	
Db	14	TCCTTTTGTTTTTT 1	

RESULT 12	AA593860/c	LOCUS	DEFINITION
	AA593860	nm19103.s1	NCI_CGAP_Col2 Homo sapiens CDNA clone IMAGE:1084349 3'
			mRNA sequence.

ACCESSION	AA593860	
VERSION	AA593860.1	GI:2408538
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 414)	NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap .	National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.

cdDNA Library Preparation: Stratagene, Inc.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrp/image/image.html
 Insert Length: 1204 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 406.

FEATURES

Source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1084349"
/sex="mixed"
/tissue_type="colon tumor"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="NCI CGAP C012"
/note="Organ: colon; Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: oligo dt. Pooled colon tumors. 5' adaptor sequence: 5' GAATTCGACACGAG 3' 3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."

```

ORIGIN

Query Match 28.7%; Score 414; DB 9; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.3e-46;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 999 CGCCGCTTCTGGGTACAGCTGGCAACCGCGCTCTTGTGCTTCTTCTCGAGGGCGCT 1058
414 CGCCGCTTCTGGGTACAGCTGGCAACCGCGCTCTTGTGCTTCTTCTCGAGGGCGCT 355

QY 1059 GGTGAGTCTCCAGTATGTTCGGCCCGACGCTCTTTCGACCCCTTTCGACCAAGCGCCAA 1118
354 GGTGAGTCTCCAGTATGTTCGGCCCGACGCTCTTTCGACCCCTTTCGACCAAGCGCCAA 295

QY 1119 GGACTGACGACGAGAGAGAGGGGGCTCAGCTTCTTATCTCTCGGACCCACTGACAGCA 1178
294 GGACTGACGACGAGAGAGAGGGGGCTCAGCTTCTTATCTCTCGGACCCACTGACAGCA 235

QY 1179 GGCCGCTCTCCAGACTTAATATGATCACTAACCTGTGAGGGGGACCAATCTGGA 1238
234 GGCCGCTCTCCAGACTTAATATGATCACTAACCTGTGAGGGGGACCAATCTGGA 175

QY 1239 CTCCTTCCCGCTTGGACATCGACGCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1298
174 CTCCTTCCCGCTTGGACATCGACGCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 115

QY 1299 GAGAGCTCCAGAGAGGAG 1358
114 GAGAGCTCCAGAGAGGAG 55

QY 1359 CAGGGAAGTCTCCTGGGGCAGTCTGTAATTAACCTTTTCTTTTCTTTT 1412
54 CAGGGAAGTCTCCTGGGGCAGTCTGTAATTAACCTTTTCTTTTCTTTT 1

Db

RESULT 13
AI983793/c 411 bp mRNA linear EST 27-OCT-1999
LOCUS WU20C09.x1 Soares_Dieckgraefe_colon_NHCD Homo sapiens cDNA clone
DEFINITION IMAGE:2520592.3', mRNA sequence.
AI983793
AI983793.1 GI:5811012
EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 411)
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 400.
Location/Qualifiers
1. 411
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2520592"
/tissue_type="colonic mucosa from 3 patients with Crohn's
disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares_Dieckgraefe_colon_NHCD"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGGAGCGCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library

FEATURES
Source

Query Match 28.4%; Score 409.4; DB 9; Length 411;
Best Local Similarity 99.8%; Pred. No. 5.6e-46;
Matches 410; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1008 CTGGGTACGCTGGCAACCGGCGCTCTGTGCTCTTCTCTCGAGGGCGCTGTGAGTCT 1067
411 CTGGGTACGCTGGCAACCGGCGCTCTGTGCTCTTCTCTCGAGGGCGCTGTGAGTCT 352

QY 1068 CCAGTATGTTGGGCCAGCGCTCTTTCGACCCCTTTCGACCAAGCGCCAGACTGCAG 1127
351 CCAGTATGTTGGGCCAGCGCTCTTTCGACCCCTTTCGACCAAGCGCCAGACTGCAG 292

QY 1128 CCAGAGAGAGGGGCTCAGCTTCTTATCTCTCGGACCCACTGCACAGAGAGAGAGAG 1187
291 CCAGAGAGAGGGGCTCAGCTTCTTATCTCTCGGACCCACTGCACAGAGAGAGAGAG 232

QY 1188 CCCAGCTTAATATGATCACTAACCTGTGAGGGGGACCAATCTGACTCTCTCC 1247
231 CCCAGCTTAATATGATCACTAACCTGTGAGGGGGACCAATCTGACTCTCTCC 172

QY 1248 CGCCTTGGGACATCGCAGGCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1307
171 CGCCTTGGGACATCGCAGGCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 112

QY 1308 AGGAGGGGACTGAGCGCTGCTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1367
111 AGGAGGGGACTGAGCGCTGCTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 52

QY 1368 TCTCCTGGGGGAGTCTGTAATTAACCTTTTCTTTTCTTTTCTTTT 1418
51 TCTCCTGGGGGAGTCTGTAATTAACCTTTTCTTTTCTTTTCTTTT 1

Db

RESULT 14
AI274929/c 405 bp mRNA linear EST 29-JAN-1999
LOCUS q149c11.x1 NCI CGAP Co8 Homo sapiens cDNA clone IMAGE:1875668.3'
DEFINITION similar to contains TARI. t3 TARI repetitive element ;, mRNA
sequence.
AI274929
AI274929.1 GI:3897203
EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 405)
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/www-bio.llnl.gov/bdtp/image/image.html at:
Insert Length: 1458 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 395.

ORIGIN
went through one round of normalization. Tissue samples
provided by Dr. Brian Dieckgraefe (Washington University,
dieck@im.wustl.edu); colonic mucosa represents a range of
disease involvement from moderate to severe Crohn's
disease; samples include both perforating (fistulas) and
non-perforating samples. Library constructed by Bento
Soares and M. Fatima Bernaldo."

FEATURES
source

Location/Qualifiers
1. .405
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1875668"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI CGAP Co8"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 28.1%; Score 405; DB 9; Length 405;
Best Local Similarity 100.0%; Pred. No. 2.2e-45;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1009 TGGGTACGCTGGCAACCGGCTCTTCTCTCGAGGGCCGTGTGAGTCTC 1068
|||||
Db 405 TGGGTACGCTGGCAACCGGCTCTTCTCTCGAGGGCCGTGTGAGTCTC 346
QY 1069 CAGTATGTTGGGCCAGCGCTCTTGGACCTTCTGACCAAGCGCCAGACTGCAGC 1128
|||||
Db 345 CAGTATGTTGGGCCAGCGCTCTTGGACCTTCTGACCAAGCGCCAGACTGCAGC 286
QY 1129 CAGGAGAGAGGGGCTCACTCTTATCTCTGCGACCACTGCACAAGCGGCTCTC 1188
|||||
Db 285 CAGGAGAGAGGGGCTCACTCTTATCTCTGCGACCACTGCACAAGCGGCTCTC 226
QY 1189 CCAGACTTAAATGTATCACTAATCTGTAGGGGACCAATCTGACTCTTCCCC 1248
|||||
Db 225 CCAGACTTAAATGTATCACTAATCTGTAGGGGACCAATCTGACTCTTCCCC 166
QY 1249 GCCTTGGACATCGCAGGCGGGAGAGAGTGCCTCCAGGCTGTGAGAGCTCCA 1308
|||||
Db 165 GCCTTGGACATCGCAGGCGGGAGAGAGTGCCTCCAGGCTGTGAGAGCTCCA 106
QY 1309 GGAAGGGCACTGAGCGCTGTGCGCGGAGGCTCTGGACATCCGACGACCAAGGAAAGT 1368
|||||
Db 105 GGAAGGGCACTGAGCGCTGTGCGCGGAGGCTCTGGACATCCGACGACCAAGGAAAGT 46
QY 1369 CTCCTGGGGCGATCTGTAAATAACCTTTTCTTTGTTT 1413
|||||
Db 45 CTCCTGGGGCGATCTGTAAATAACCTTTTCTTTGTTT 1

RESULT 15
AI346155/c 405 bp mRNA linear EST 02-FEB-1999
LOCUS qp43f12.x1 NCI CGAP Co8 Homo sapiens cDNA clone IMAGE:1925807 3'
DEFINITION similar to contains TAX1.ct3 TAR1 repetitive element ;, mRNA
sequence.

ACCESSION AI346155
VERSION AI346155.1 GI:4083361
KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 405)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

FEATURES
source

Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www.bio.lnl.gov/bbrp/image/image.html
Insert Length: 1488 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 375.

Location/Qualifiers

1. .405
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1925807"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI CGAP Co8"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 27.8%; Score 400.2; DB 9; Length 405;
Best Local Similarity 99.3%; Pred. No. 9.9e-45;
Matches 402; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1009 TGGGTACGCTGGCAACCGGCTCTTCTCTCGAGGGCCGTGTGAGTCTC 1068
|||||
Db 405 TGGGTACGCTGGCAACCGGCTCTTCTCTCGAGGGCCGTGTGAGTCTC 346
QY 1069 CAGTATGTTGGGCCAGCGCTCTTGGACCTTCTGACCAAGCGCCAGACTGCAGC 1128
|||||
Db 345 CAGTATGTTGGGCCAGCGCTCTTGGACCTTCTGACCAAGCGCCAGACTGCAGC 286
QY 1129 CAGGAGAGAGGGGCTCACTCTTATCTCTGCGACCACTGCACAAGCGGCTCTC 1188
|||||
Db 285 CAGGAGAGAGGGGCTCACTCTTATCTCTGCGACCACTGCACAAGCGGCTCTC 226
QY 1189 CCAGACTTAAATGTATCACTAATCTGTAGGGGACCAATCTGACTCTTCCCC 1248
|||||
Db 225 CCAGACTTAAATGTATCACTAATCTGTAGGGGACCAATCTGACTCTTCCCC 166
QY 1249 GCCTTGGACATCGCAGGCGGGAGAGAGTGCCTCCAGGCTGTGAGAGCTCCA 1308
|||||
Db 165 GCCTTGGACATCGCAGGCGGGAGAGAGTGCCTCCAGGCTGTGAGAGCTCCA 106
QY 1309 GGAAGGGCACTGAGCGCTGTGCGCGGAGAGGCTCTGGACATCCGACGACCAAGGAAAGT 1368
|||||
Db 105 GGAAGGGCACTGAGCGCTGTGCGCGGAGAGGCTCTGGACATCCGACGACCAAGGAAAGT 46
QY 1369 CTCCTGGGGCGATCTGTAAATAACCTTTTCTTTGTTT 1413
|||||
Db 45 CTCCTGGGGCGATCAAAAAATAACCTTTTCTTTGTTT 1

Search completed: February 23, 2004, 23:20:16
Job time : 3671.04 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:21:32 ; Search time 20 Seconds
(without alignments)
1539.064 Million cell updates/sec

Title: US-09-936-456-2

Perfect score: 320

Sequence: 1 MTLNMGVLPFYPPQPRHAAGF.....GDP LHKQALPDLCITTNL 320

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: PIR 78:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	2.5	141	2 AG0657	hypothetical perip
2	8	2.5	227	2 B83505	hypothetical prote
3	8	2.5	253	2 S49055	thioesterase - Str
4	8	2.5	417	2 JC7092	PsuI protein - fis
5	8	2.5	575	2 B83586	probable permease
6	8	2.5	615	2 B91272	enzyme in methyl-d
7	8	2.5	615	2 B86113	enzyme in methyl-d
8	8	2.5	615	2 PH0853	methyl-directed mi
9	8	2.5	618	2 A33588	mismatch repair pr
10	8	2.5	618	2 AG1048	DNA mismatch repair
11	8	2.5	629	2 E64046	mismatch repair pr
12	8	2.5	633	2 F83028	DNA mismatch repai
13	8	2.5	635	2 AC0046	DNA mismatch repai
14	8	2.5	653	2 A82334	DNA mismatch repai
15	8	2.5	669	2 C89904	DNA mismatch repai
16	8	2.5	15	2 I46909	voltage-dependent
17	7	2.2	86	2 D91094	type III secretion
18	7	2.2	86	2 H85939	type III secretion
19	7	2.2	93	2 F95340	hypothetical prote
20	7	2.2	102	2 H72619	probable NADH-plas
21	7	2.2	116	2 C86120	probable growth in
22	7	2.2	116	2 D49339	cell growth regula
23	7	2.2	116	2 C91279	lipid transfer pro
24	7	2.2	118	2 T02049	hypothetical prote
25	7	2.2	126	2 F83294	hypothetical prote
26	7	2.2	127	2 F83803	hypothetical prote
27	7	2.2	131	1 RDEB15	hypothetical prote
28	7	2.2	182	2 S69469	hypothetical prote
29	7	2.2	205	1 F71193	hypothetical prote

30	7	2.2	205	2 G70338	conserved hypothet
31	7	2.2	206	2 D70935	hypothetical prote
32	7	2.2	209	2 E75152	hypothetical prote
33	7	2.2	210	2 A87022	conserved hypothet
34	7	2.2	215	2 E83564	hypothetical prote
35	7	2.2	218	2 C83152	hypothetical prote
36	7	2.2	243	2 C69754	toxic cation resis
37	7	2.2	249	2 H84688	probable thaumat
38	7	2.2	258	2 F90029	hypothetical prote
39	7	2.2	263	2 D95180	glutathione S-tran
40	7	2.2	263	2 H98047	conserved hypothet
41	7	2.2	266	2 E64955	fliv protein precu
42	7	2.2	266	2 C90961	probable periplasm
43	7	2.2	266	2 C85809	probable periplasm
44	7	2.2	266	2 AD0750	cysteine-binding pe
45	7	2.2	266	2 T34172	hypothetical prote

ALIGNMENTS

RESULT 1

AG0657

hypothetical periplasmic protein STY1364 [imported] - Salmonella enterica subsp. ente

C/Species: Salmonella enterica subsp. enterica serovar Typhi

A/Note: this species has also been called Salmonella typhi

C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C/Accession: AG0657

R/ParKhl1, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se

A/Reference number: AB0502; MUID:21534947; PMID:11677608

A/Accession: AG0657

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-141 <PAR>

A/Cross-references: GB:AL513382; PIDN:CAD01632.1; PID:G16502485; GSPDB:GN00176

C/Genetics:

A/Gene: STY1364

Query Match

Best Local Similarity 2.5%; Score 8; DB 2; Length 141;

Matches 8; Conservative 100.0%; Pred. No. 5.1;

0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230

Db 11 ALASISSV 18

RESULT 2

B83505

hypothetical protein PA118 [imported] - Pseudomonas aeruginosa (strain PA01)

C/Species: Pseudomonas aeruginosa

C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C/Accession: B83505

R/Stover, C.K.; Pham, X.C.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

..; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A/Reference number: A82950; MUID:20437337; PMID:10984043

A/Accession: B83505

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-227 <STO>

A/Cross-references: GB:AE004542; GB:AE004091; NID:G9947034; PIDN:AA04507.1; GSPDB:GN

C/Genetics:

A/Gene: PA118

Query Match 2.5%; Score 8; DB 2; Length 227;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 RVGLVGL 105
|||
Db 105 RVGLVGL 112

RESULT 3

S49055
thioesterase - Streptomyces fradiae (strain T59235)

C/Species: Streptomyces fradiae

A/Variety: strain T59235

C/Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jul-2003

C/Accession: S49055

R/Merson-Davies, L.A.; Cundliffe, E.

Mol. Microbiol. 13, 349-355, 1994

A/Title: Analysis of five tylosin biosynthetic genes from the tytIBA region of the Strept

A/Reference number: S49051; MUID:95075319; PMID:7984112

A/Accession: S49055

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-253 <MER>

A/Cross-references: EMBL:U08223; NID:G6849140; PIDN:AAA21345.1; PID:G473601

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994

C/Superfamily: thioesterase, type II; oleoyl-[acyl-carrier-protein] hydrolase homology
F.16-232/Domain: oleoyl-[acyl-carrier-protein] hydrolase homology <ACPH>

Query Match 2.5%; Score 8; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 LTTGAFAL 218
|||
Db 213 LTTGAFAL 220

RESULT 4

UC7092

Psul protein - fission yeast (Schizosaccharomyces pombe)

C/Species: Schizosaccharomyces pombe

C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000

C/Accession: UC7092

R/Oml, K.; Sonoda, H.; Nagata, K.; Sugita, K.

Biochem. Biophys. Res. Commun. 262, 368-374, 1999

A/Title: Cloning and characterization of psul+, a new essential fission yeast gene invol

A/Reference number: UC7092; MUID:99393312; PMID:10462482

A/Accession: UC7092

A/Molecule type: DNA

A/Residues: 1-417 <OMI>

A/Cross-references: DDBJ:AB009980; NID:G5822708; PIDN:BAAB3907.1; PID:G5822709

C/Genetics:

A/Gene: psul+

C/Superfamily: Saccharomyces NCAs protein

C/Keywords: cell wall synthesis

Query Match 2.5%; Score 8; DB 2; Length 417;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 LALLTTGA 215
|||
Db 9 LALLTTGA 16

RESULT 5

B83586

probable permease PA0476 [imported] - Pseudomonas aeruginosa (strain PA01)

C/Species: Pseudomonas aeruginosa

C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C/Accession: B83586

R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.;
.i. Lofy, S.; Olson, M.V.

Nature 406, 959-964, 2000

A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic p

A/Reference number: A82950; MUID:20437337; PMID:10984043

A/Accession: B83586

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-575 <STO>

A/Cross-references: GB:AE004485; GB:AE004091; NID:G9946332; PIDN:AA03865.1; GSPDB:G

A/Experimental source: strain PA01

C/Genetics:

A/Gene: PA0476

Query Match 2.5%; Score 8; DB 2; Length 575;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 GAAPFWTL 254
|||
Db 316 GAAPFWTL 323

RESULT 6

B91272

enzyme in methyl-directed mismatch repair [imported] - Escherichia coli (strain O157

C/Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

C/Accession: B91272

R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han,

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and

A/Reference number: A99629; MUID:21156231; PMID:11258796

A/Accession: B91272

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-615 <HAY>

A/Cross-references: GB:BA000007; PIDN:BA838569.1; PID:G1384623; GSPDB:GN00154

A/Experimental source: strain O157:H7, substrain RIMD 0509952

C/Genetics:

A/Gene: ECS5146

C/Superfamily: mismatch repair protein hexB

Query Match 2.5%; Score 8; DB 2; Length 615;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
|||
Db 98 ALASISSV 105

RESULT 7

B86113

enzyme in methyl-directed mismatch repair [imported] - Escherichia coli (strain O157

C/Species: Escherichia coli

C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C/Accession: B86113

R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Ma

iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apod

Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A/Reference number: A85480; MUID:21074935; PMID:11206551

A/Accession: B86113

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-615 <STO>

A/Cross-references: GB:AE005174; NID:G12519155; PIDN:AA059366.1; GSPDB:GN00145; UWGP

A/Experimental source: strain O157:H7, substrain EDL933

C/Genetics:

A/Gene: mutL

C/Superfamily: mismatch repair protein hexB

Query Match 2.5%; Score 8; DB 2; Length 615;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 98 ALASISSV 105

RESULT 8
PH0853
methyl-directed mismatch repair protein mutL - Escherichia coli (strain K-12)
C/Species: Escherichia coli
C/Date: 16-Sep-1992 #sequence revision 16-Sep-1992 #text change 01-Mar-2002
C/Accession: PH0853; S56395; A37318; S40056; E65227; S23011
R/Tsui, H.C.T.; Mandavilli, B.S.; Winkler, M.E.
Nucleic Acids Res. 20, 2379, 1992
A/Title: Nonconserved segment of the mutL protein from Escherichia coli K-12 and Salmons
A/Reference number: PH0853; MUID:92279041; PMID:1594459
A/Accession: PH0853
A/Molecule type: DNA
A/Residues: 1-615 <TSU>
A/Cross-references: EMBL:Z11831; NID:g42066; PIDN:CAA77850.1; PID:g42067
A/Experimental source: strain K12
A/Note: the complete nucleotide sequence is not shown; the complete translation is not s
A/Note: translation of the nucleotide sequence is not complete
R/Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A/Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.
A/Reference number: S56314; MUID:95334362; PMID:7610040
A/Accession: S56395
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-615 <BUR>
A/Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97066.1; PID:g537011
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
R/Connolly, D.M.; Winkler, M.E.
J. Bacteriol. 173, 1711-1721, 1991
A/Title: Structure of Escherichia coli K-12 miaA and characterization of the mutator phe
A/Reference number: A37318; MUID:91154127; PMID:1999389
A/Accession: A37318
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 596-615 <CON>
A/Cross-references: GB:M63655; GB:M37459; NID:g146858; PIDN:AAA24173.1; PID:g146859
R/Tsui, H.C.T.; Zhao, G.; Peng, G.; Leung, H.C.E.; Winkler, M.E.
Mol. Microbiol. 11, 189-202, 1994
A/Title: The mutL repair gene of Escherichia coli K-12 forms a superoperon with a gene e
A/Reference number: S40053; MUID:94195106; PMID:7511774
A/Accession: S40056
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-23 <TSW>
A/Cross-references: EMBL:L19346; NID:g304911; PIDN:AAA20098.1; PID:g304915
R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cd
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MUID:97426617; PMID:9278503
A/Accession: E65227
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-615 <BLAT>
A/Cross-references: GB:AE000489; GB:U00096; NID:g1790607; PIDN:AAC77127.1; PID:g1790612;
A/Experimental source: strain K-12, substrain MG1655
C/Genetics:
A/Gene: mutL
A/Map position: 95 min
C/Function:
A/Description: this protein is required for methyl-directed mismatch repair in vivo and
C/Superfamily: mismatch repair protein hexB
C/Keywords: DNA repair

Query Match 2.5%; Score 8; DB 2; Length 615;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 98 ALASISSV 105

RESULT 9
A33588
mismatch repair protein mutL - Salmonella typhimurium
C/Species: Salmonella typhimurium
C/Date: 27-Feb-1990 #sequence revision 27-Feb-1990 #text change 08-Oct-1999
C/Accession: A33588
R/Mankovich, J.A.; McIntyre, C.A.; Walker, G.C.
J. Bacteriol. 171, 5325-5331, 1989
A/Title: Nucleotide sequence of the Salmonella typhimurium mutL gene required for mis
ae.
A/Reference number: A33588; MUID:90008766; PMID:2676972
A/Accession: A33588
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-618 <MAN>
A/Cross-references: GB:M9687; NID:g154185; PIDN:AAA27166.1; PID:g154186
C/Superfamily: mismatch repair protein hexB

Query Match 2.5%; Score 8; DB 2; Length 618;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 98 ALASISSV 105

RESULT 10
AG1048
DNA mismatch repair protein [imported] - Salmonella enterica subsp. enterica serovar
C/Species: Salmonella enterica subsp. enterica serovar Typhi
A/Note: this species has also been called Salmonella typhi
C/Date: 09-Nov-2001 #sequence revision 09-Nov-2001 #text change 18-Nov-2002
C/Accession: AG1048
R/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A/Reference number: AB0502; MUID:21534947; PMID:11677608
A/Accession: AG1048
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-618 <PAR>
A/Cross-references: GB:AL513382; PIDN:CAD06836.1; PID:g16505486; GSPDB:GN00176
C/Genetics:
A/Gene: mutL
C/Superfamily: mismatch repair protein hexB

Query Match 2.5%; Score 8; DB 2; Length 618;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 98 ALASISSV 105

RESULT 11
B64046
mismatch repair protein mutL - Haemophilus influenzae (strain Rd KW20)
C/Species: Haemophilus influenzae

C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C/Accession: B64046
R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A/Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A/Reference number: A64000; MUID:95350630; PMID:7542800
A/Accession: E64046
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-629 <TIGR>
A/Cross-references: GB:U32692; GB:L42023; NID:G1573013; PIDN:AAC21745.1; PID:G1573016; T
C/Genetics:
A/Gene: mutL
C/Superfamily: mismatch repair protein hexB
C/Keywords: DNA repair

Query Match 2.5%; Score 8; DB 2; Length 629;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
|||||
Db 98 ALASISSV 105

RESULT 12

F83028
DNA mismatch repair protein MutL PA4946 [imported] - Pseudomonas aeruginosa (strain PA01
C/Species: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: F83028
R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A/Reference number: A82950; MUID:20437337; PMID:10984043
A/Accession: F83028
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-633 <STO>
A/Cross-references: GB:AE004907; GB:AE004091; NID:G9951217; PIDN:AG08331.1; GSPDB:GN001
A/Experimental source: strain PA01
C/Genetics:
A/Gene: mutL; PA4946
C/Superfamily: mismatch repair protein hexB

Query Match 2.5%; Score 8; DB 2; Length 633;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
|||||
Db 102 ALASISSV 109

RESULT 13

AC0046
DNA mismatch repair protein [imported] - Yersinia pestis (strain CO92)
C/Species: Yersinia pestis
C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C/Accession: AC0046
R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,
Nature 413, 523-527, 2001
A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A/Reference number: AB0001; MUID:21470413; PMID:11586360
A/Accession: AC0046
A/Status: preliminary

A/Molecule type: DNA
A/Residues: 1-635 <KUR>
A/Cross-references: GB:AL590842; PIDN:CAC89230.1; PID:G15978469; GSPDB:GN00175
C/Genetics:
A/Gene: mutL
C/Superfamily: mismatch repair protein hexB

Query Match 2.5%; Score 8; DB 2; Length 635;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
|||||
Db 98 ALASISSV 105

RESULT 14

A82334
DNA mismatch repair protein MutL VC0345 [imported] - Vibrio cholerae (strain N16961
C/Species: Vibrio cholerae
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C/Accession: A82334
R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R
chardson, D.; Ermolaeva, M.D.; Yamathavan, J.; Bass, S.; Qin, H.; Dragoi, I.; Seller
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A/Reference number: A82035; MUID:20406833; PMID:10952301
A/Accession: A82334
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-653 <HEI>
A/Cross-references: GB:AE004123; GB:AE003852; NID:G9654756; PIDN:AAF93518.1; GSPDB:G
A/Experimental source: serogroup O1; strain N16961; biotype El Tor
C/Genetics:
A/Gene: VC0345
A/Map position: 1
C/Superfamily: mismatch repair protein hexB

Query Match 2.5%; Score 8; DB 2; Length 653;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
|||||
Db 98 ALASISSV 105

RESULT 15

C89904
DNA mismatch repair protein [imported] - Staphylococcus aureus (strain N315)
C/Species: Staphylococcus aureus
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C/Accession: C89904
R/Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.;
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu,
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A/Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A/Reference number: A89758; MUID:21311952; PMID:11418146
A/Accession: C89904
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-669 <KUR>
A/Cross-references: GB:BA000018; PID:G13701096; PIDN:BAB42391.1; GSPDB:GN00149
A/Experimental source: strain N315
C/Genetics:
A/Gene: mutL
C/Superfamily: mismatch repair protein hexB

Query Match 2.5%; Score 8; DB 2; Length 669;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	223	ALASISSV	230
Db	99	ALASISSV	106

Search completed: February 23, 2004, 19:35:20
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:32:12 ; Search time 23 Seconds

(without alignments)
718.274 Million cell updates/sec

Title: US-09-936-456-2

Perfect score: 320

Sequence: 1 MTLWNGVLPFYQPRHAGF.....GDP LHKQALPDLKCTTNL 320

Scoring table: OLIGO

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	2.5	227	4	US-09-252-991A-19468
2	8	2.5	326	4	US-09-134-001C-4934
3	8	2.5	361	2	US-08-209-521-5
4	8	2.5	361	3	US-08-961-810-1
5	8	2.5	361	3	US-08-352-902D-1
6	8	2.5	361	4	US-09-265-503B-1
7	8	2.5	602	4	US-09-252-991A-22527
8	8	2.5	615	3	US-08-676-444-4
9	8	2.5	659	4	US-09-252-991A-33126
10	8	2.5	674	4	US-09-543-681A-5715
11	7	2.2	91	4	US-09-369-247-83
12	7	2.2	118	1	US-08-207-904-4
13	7	2.2	116	1	US-08-207-904-19
14	7	2.2	142	4	US-09-489-039A-9418
15	7	2.2	144	4	US-09-252-991A-31261
16	7	2.2	152	4	US-09-543-681A-6097
17	7	2.2	158	4	US-09-489-039A-8515
18	7	2.2	169	4	US-09-328-352-6931
19	7	2.2	225	4	US-09-465-901-18
20	7	2.2	226	4	US-09-252-991A-27390
21	7	2.2	246	3	US-08-896-933-33
22	7	2.2	246	4	US-09-314-235-33
23	7	2.2	262	3	US-08-948-265-2
24	7	2.2	262	3	US-08-948-265-4
25	7	2.2	263	2	US-08-474-379C-42
26	7	2.2	263	3	US-09-146-249A-42
27	7	2.2	263	3	US-08-206-188B-42

28	7	2.2	264	4	US-09-252-991A-21359	Sequence 21359, A
29	7	2.2	265	1	US-07-688-352C-42	Sequence 42, Appl
30	7	2.2	265	4	US-09-252-991A-32803	Sequence 32803, A
31	7	2.2	276	4	US-09-543-681A-4338	Sequence 4338, Ap
32	7	2.2	287	4	US-09-328-352-6652	Sequence 6652, Ap
33	7	2.2	299	4	US-09-393-634-35	Sequence 35, Appl
34	7	2.2	300	4	US-09-489-039A-12250	Sequence 12250, A
35	7	2.2	311	4	US-09-134-001C-3846	Sequence 3846, Ap
36	7	2.2	313	3	US-08-926-842B-62	Sequence 62, Appl
37	7	2.2	315	4	US-09-489-039A-9274	Sequence 9274, Ap
38	7	2.2	336	2	US-08-474-379C-36	Sequence 36, Appl
39	7	2.2	336	3	US-09-146-249A-36	Sequence 36, Appl
40	7	2.2	336	3	US-08-206-188B-36	Sequence 151, App
41	7	2.2	341	4	US-08-311-731A-151	Sequence 44, Appl
42	7	2.2	386	1	US-07-688-352C-44	Sequence 44, Appl
43	7	2.2	386	5	PCT-US91-02714-41	Sequence 44, Appl
44	7	2.2	404	2	US-08-474-379C-44	Sequence 44, Appl
45	7	2.2	404	3	US-09-146-249A-44	Sequence 44, Appl

ALIGNMENTS

```
RESULT 1
US-09-252-991A-19468
; Sequence 19468, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252, 991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074, 788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094, 190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19468
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19468

Query Match      2.5% Score 8; DB 4; Length 227;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      98 RVGLLVGL 105
Db      105 RVGLLVGL 112

RESULT 2
US-09-134-001C-4934
; Sequence 4934, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134, 001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064, 964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055, 779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4934
; LENGTH: 326
; TYPE: PRT
```

ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4934

Query Match 2.5%; Score 8; DB 4; Length 326;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 100 ALASISSV 107

RESULT 3

US-08-209-521-5
Sequence 5, Application US/08209521
Patent No. 5922855

GENERAL INFORMATION:

APPLICANT: Liskay, Robert M.
APPLICANT: Bronner, C. Eric
APPLICANT: Baker, Sean M.
APPLICANT: Bollag, Roni J.
APPLICANT: Kolodner, Richard D.
TITLE OF INVENTION: MAMMALIAN DNA MISMATCH REPAIR GENES
TITLE OF INVENTION: MMLH1 AND HPM51
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kolisch, Hartwell, Dickinson, McCormack &
ADDRESSEE: Heuser
STREET: 520 S.W. Yamhill, Suite 200
CITY: Portland
STATE: Oregon
COUNTRY: US
ZIP: 97204

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/209,521
FILING DATE: 08-MAR-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Van Rysselberghe, Pierre C.
REGISTRATION NUMBER: 33,557
REFERENCE/DOCKET NUMBER: OHSU 306A

TELECOMMUNICATION INFORMATION:
TELEPHONE: (503) 224-6655
TELEFAX: (503) 295-6679

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-209-521-5

Query Match 2.5%; Score 8; DB 2; Length 361;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 98 ALASISSV 105

RESULT 4

US-08-961-810-1
Sequence 1, Application US/08961810
Patent No. 6165713

GENERAL INFORMATION:
APPLICANT: Liskay, Robert M.

APPLICANT: Bronner, C. Eric
APPLICANT: Baker, Sean M.
APPLICANT: Bollag, Roni J.

APPLICANT: Kolodner, Richard D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS RELATING TO DNA
TITLE OF INVENTION: MISMATCH REPAIR GENES
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kolisch, Hartwell, Dickinson, McCormack &
ADDRESSEE: Heuser
STREET: 520 S.W. Yamhill Street, Suite 200
CITY: Portland
STATE: Oregon
COUNTRY: U.S.A.
ZIP: 97204

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,810
FILING DATE:
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Van Rysselberghe, Pierre C.
REGISTRATION NUMBER: 33,557
REFERENCE/DOCKET NUMBER: OHSU 306B

TELECOMMUNICATION INFORMATION:
TELEPHONE: (503) 224-6655
TELEFAX: (503) 295-6679
TELEX: 360619

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 361 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA

US-08-961-810-1

Query Match 2.5%; Score 8; DB 3; Length 361;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 98 ALASISSV 105

RESULT 5

US-08-352-902D-1
Sequence 1, Application US/08352902D
Patent No. 6191268

GENERAL INFORMATION:

APPLICANT: Liskay, Robert M.
APPLICANT: Bronner, C. Eric
APPLICANT: Baker, Sean M.
APPLICANT: Bollag, Roni J.
APPLICANT: Kolodner, Richard D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS RELATING TO DNA
TITLE OF INVENTION: MISMATCH REPAIR GENES
NUMBER OF SEQUENCES: 149
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kolisch, Hartwell, Dickinson, McCormack &
ADDRESSEE: Heuser
STREET: 520 S.W. Yamhill Street, Suite 200
CITY: Portland
STATE: Oregon
COUNTRY: U.S.A.
ZIP: 97204

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/352, 902D
FILING DATE: 09-Dec-1994
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Van Rysselberghe, Pierre C.
REGISTRATION NUMBER: 33,557
REFERENCE/DOCKET NUMBER: OHSU 306B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (503) 224-6655
TELEFAX: (503) 295-6679
TELEX: 360619
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-08-352-902D-1

Query Match 2.5%; Score 8; DB 3; Length 361;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 98 ALASISSV 105

RESULT 6
US-09-265-503B-1
Sequence 1, Application US/09265503B
Patent No. 6538108
GENERAL INFORMATION:
APPLICANT: Liskay, Robert M.
APPLICANT: Bronner, C. Eric
APPLICANT: Baker, Sean M.
APPLICANT: Bollag, Roni J.
APPLICANT: Kolodner, Richard D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS
TITLE OF INVENTION: RELATING TO DNA MISMATCH REPAIR GENES
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kolisch, Hartwell, Dickinson, McCormack & Heuser
STREET: 520 S.W. Yamhill Street, Suite 200
CITY: Portland
STATE: Oregon
COUNTRY: U.S.A.
ZIP: 97204
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/265,503B
FILING DATE: March 10, 1999
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Van Rysselberghe, Pierre C.
REGISTRATION NUMBER: 33,557
REFERENCE/DOCKET NUMBER: OHSU 306D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (503) 224-6655
TELEFAX: (503) 295-6679
TELEX: 360619
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 361 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-265-503B-1

Query Match 2.5%; Score 8; DB 4; Length 361;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 98 ALASISSV 105

RESULT 7
US-09-252-991A-22527
Sequence 22527, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 22527
LENGTH: 602
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22527

Query Match 2.5%; Score 8; DB 4; Length 602;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 GAAPWVTL 254
Db 343 GAAPWVTL 350

RESULT 8
US-08-676-444-44
Sequence 44, Application US/08676444A
Patent No. 6294325
GENERAL INFORMATION:
APPLICANT: Wetmur, James G.
TITLE OF INVENTION: CLONING AND EXPRESSION OF THERMOSTABLE
TITLE OF INVENTION: MUTL GENES AND PROTEINS AND USES THEREFOR
FILE REFERENCE: MSM95-02
CURRENT APPLICATION NUMBER: US/08/676,444A
CURRENT FILING DATE: 1996-07-05
NUMBER OF SEQ ID NOS: 48
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 44
LENGTH: 615
TYPE: PRT
ORGANISM: Escherichia coli
US-08-676-444-44

Query Match 2.5%; Score 8; DB 3; Length 615;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 98 ALASISSV 105

RESULT 9
US-09-252-991A-33126
; Sequence 33126, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 33126
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-33126

Query Match 2.5%; Score 8; DB 4; Length 659;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 223 ALASISSV 230
Db 128 ALASISSV 135

RESULT 10
US-09-543-681A-5715
; Sequence 5715, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETTON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5715
; LENGTH: 674
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5715

Query Match 2.5%; Score 8; DB 4; Length 674;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 223 ALASISSV 230
Db 103 ALASISSV 110

RESULT 11
US-09-369-247-83
; Sequence 83, Application US/09369247
; Patent No. 6569992
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 44 Human Secreted Proteins
; FILE REFERENCE: P2024P1
; CURRENT APPLICATION NUMBER: US/09/369,247
; CURRENT FILING DATE: 1999-08-05
; EARLIER APPLICATION NUMBER: 60/074,118

; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/074,157
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/074,137
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/074,341
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/074,141
; EARLIER FILING DATE: 1998-02-09
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 83
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (91)
; OTHER INFORMATION: Xaa equals stop translation
US-09-369-247-83

Query Match 2.2%; Score 7; DB 4; Length 91;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 29 LVFLALA 35
Db 5 LVFLALA 11

RESULT 12
US-08-207-904-4
; Sequence 4, Application US/08207904
; Patent No. 5477002
; GENERAL INFORMATION:
; APPLICANT: Tuttle, AnnMarie
; TITLE OF INVENTION: Antier-Specific cDNA Sequences, Genomic
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/207,904
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/908,242
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lazar, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: CGC 1624
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8615
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 118 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-207-904-4

Query Match 2.2%; Score 7; DB 1; Length 118;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 LVFLALA 35
|||
Db 5 LVFLALA 11

RESULT 13
US-08-207-904-19
Sequence 19, Application US/08207904
Patent No. 5477002

GENERAL INFORMATION:
APPLICANT: Tuttle, AnnMarie
APPLICANT: Crossland, Lyle D.
TITLE OF INVENTION: Another-Specific cDNA Sequences, Genomic
TITLE OF INVENTION: DNA Sequences and Recombinant DNA Sequences
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/207,904
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/908,242
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Lazar, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: CGC 1624
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8615
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 118 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-207-904-19

Query Match 2.2%; Score 7; DB 1; Length 118;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 LVFLALA 35
|||
Db 5 LVFLALA 11

RESULT 14
US-09-489-039A-9418

Sequence 9418, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709,2004001
CURRENT APPLICATION NUMBER: US/09/489,039A

CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 9418
LENGTH: 142
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9418

Query Match 2.2%; Score 7; DB 4; Length 142;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 VLLSLFI 63
|||
Db 67 VLLSLFI 73

RESULT 15

US-09-252-991A-31261
Sequence 31261, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 31261
LENGTH: 144
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31261

Query Match 2.2%; Score 7; DB 4; Length 144;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 FSARAVT 96
|||
Db 106 FSARAVT 112

Search completed: February 23, 2004, 19:35:55
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model1

Run on: February 23, 2004, 19:33:22 ; Search time 36 Seconds

(Without alignments)
1876.918 Million cell updates/sec

Title: US-09-936-456-2

Perfect score: 320

Sequence: 1 MTLWNGVLPFYQPRHAGF.....GDP LHKQALPDLKCTITNL 320

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 809742 seqs, 211153259 residues

Word size : 0

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	320	100.0	320	US-10-187-657-1	Sequence 1, Appli
2	219	68.4	320	US-10-264-237-2502	Sequence 2502, Ap
3	19	5.9	98	US-10-106-698-7554	Sequence 7554, Ap
4	19	5.9	112	US-09-759-130B-427	Sequence 427, App
5	19	5.9	112	US-10-042-431-57	Sequence 57, Appl
6	19	5.9	298	US-10-094-749-2834	Sequence 2834, Ap
7	19	5.9	343	US-09-759-130B-425	Sequence 425, App
8	19	5.9	343	US-10-042-431-55	Sequence 55, Appl
9	8	2.5	49	US-09-764-891-2983	Sequence 2983, Ap
10	8	2.5	120	US-10-156-761-12862	Sequence 12862, A
11	8	2.5	345	US-10-128-714-8106	Sequence 8106, Ap
12	8	2.5	615	US-10-109-791A-20	Sequence 20, Appl
13	8	2.5	629	US-09-815-242-10961	Sequence 10961, A
14	8	2.5	633	US-09-815-242-12077	Sequence 12077, A
15	8	2.5	669	US-09-815-242-5262	Sequence 5262, Ap

16	8	2.5	669	9	US-09-815-242-12334	Sequence 12334, A
17	8	2.5	669	14	US-10-109-791A-19	Sequence 19, Appl
18	8	2.5	734	9	US-09-738-626-4227	Sequence 4227, Ap
19	8	2.5	772	9	US-09-935-799A-2	Sequence 2, Appli
20	8	2.5	772	9	US-09-935-799A-5	Sequence 5, Appli
21	7	2.2	20	10	US-09-986-480-234	Sequence 234, App
22	7	2.2	33	14	US-10-081-816-89	Sequence 89, Appl
23	7	2.2	54	9	US-09-729-674-40	Sequence 40, Appl
24	7	2.2	65	9	US-09-925-301-1675	Sequence 1675, Ap
25	7	2.2	87	9	US-09-864-761-44598	Sequence 44598, A
26	7	2.2	87	11	US-09-864-408A-5048	Sequence 5048, Ap
27	7	2.2	91	14	US-10-062-548-83	Sequence 83, Appl
28	7	2.2	109	9	US-09-864-761-42556	Sequence 42556, A
29	7	2.2	126	10	US-09-820-843A-91	Sequence 91, Appl
30	7	2.2	147	15	US-10-108-260A-3104	Sequence 3104, Ap
31	7	2.2	194	15	US-10-108-260A-4172	Sequence 4172, Ap
32	7	2.2	206	14	US-10-080-170-513	Sequence 513, App
33	7	2.2	209	14	US-10-156-761-10982	Sequence 10982, A
34	7	2.2	210	14	US-10-080-170-99	Sequence 99, Appl
35	7	2.2	220	15	US-10-369-493-12396	Sequence 12396, A
36	7	2.2	225	14	US-10-259-430-18	Sequence 18, Appl
37	7	2.2	225	14	US-10-259-423-18	Sequence 18, Appl
38	7	2.2	262	9	US-09-770-075-2	Sequence 2, Appli
39	7	2.2	262	9	US-09-770-075-4	Sequence 4, Appli
40	7	2.2	262	9	US-09-769-997-2	Sequence 2, Appli
41	7	2.2	262	9	US-09-769-997-4	Sequence 4, Appli
42	7	2.2	262	14	US-10-156-761-14181	Sequence 14181, A
43	7	2.2	263	10	US-09-769-787-68	Sequence 68, Appl
44	7	2.2	266	9	US-09-815-242-10197	Sequence 10197, A
45	7	2.2	282	9	US-09-815-242-14057	Sequence 14057, A

ALIGNMENTS

RESULT 1
US-10-187-657-1
Sequence 1, Application US/10187657
GENERAL INFORMATION:
APPLICANT: Lasek, Amy K.W.
APPLICANT: Baughn, Mariah R.
APPLICANT: Azimzal, Yalda
TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
FILE REFERENCE: PV-0009 CIP
CURRENT APPLICATION NUMBER: US/10/187,657
PRIOR FILING DATE: 2002-07-01
PRIOR APPLICATION NUMBER: PCT/US00/07817
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/139,565
PRIOR FILING DATE: 1999-06-16
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PERL Program
SEQ ID NO 1
LENGTH: 320
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Inocyte ID No. US20030068311A1 4901066CD1
US-10-187-657-1

Query Match 100.0%; Score 320; DB 14; Length 320;
Best Local Similarity 100.0%; Pred. No. 4.8e-293;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy	1	MTLWNGVLPFYQPRHAGFVPLTIVLFLAASFLILPGIRGSRFWLVRVLLS	60
Db	1	MTLWNGVLPFYQPRHAGFVPLTIVLFLAASFLILPGIRGSRFWLVRVLLS	60
Cy	61	LFIGAEIVAVHPSAEWFGVTNTNTSYKAFSAARVTARVGLVGLGGINITLTGTPVHQL	120
Db	61	LFIGAEIVAVHPSAEWFGVTNTNTSYKAFSAARVTARVGLVGLGGINITLTGTPVHQL	120

QY 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAEKFTPSSPCGLYHQYHLAGHYA 180
DB 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAEKFTPSSPCGLYHQYHLAGHYA 180
QY 181 SATLWVAFCFWLLSNVLLSTPAPLYGGLALLTTGAFALFGVAFALASISSVPLCPLRLGSS 240
DB 181 SATLWVAFCFWLLSNVLLSTPAPLYGGLALLTTGAFALFGVAFALASISSVPLCPLRLGSS 240
QY 241 ALTTQGAAFWVTIATGVLCLFLGGAVVSLOQYVRPSALRTLLDQSAKDCSQERGSSPLIL 300
DB 241 ALTTQGAAFWVTIATGVLCLFLGGAVVSLOQYVRPSALRTLLDQSAKDCSQERGSSPLIL 300
QY 301 GDPLHKQAAALPDLKCITTNL 320
DB 301 GDPLHKQAAALPDLKCITTNL 320

RESULT 2

US-10-264-237-2502
; Sequence 2502, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA3131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; PRIOR FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2502
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (106)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-237-2502

Query Match 68.4%; Score 219; DB 15; Length 320;
Best Local Similarity 99.7%; Pred. No. 7.2e-198;
Matches 319; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTLWNGVLPFYPPQPRHAGFSVPLLYLVFLALAASFLLLPGIRGHSRWFWLVRLLS 60
DB 1 MTLWNGVLPFYPPQPRHAGFSVPLLYLVFLALAASFLLLPGIRGHSRWFWLVRLLS 60
QY 61 LFIGAIVAVHFSAEWFGVTNTNTSYKAFSARVTVARVGLVGLGGINITLLTGPVHQL 120
DB 61 LFIGAIVAVHFSAEWFGVTNTNTSYKAFSARVTVARVGLVGLGGINITLLTGPVHQL 120
QY 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAEKFTPSSPCGLYHQYHLAGHYA 180
DB 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAEKFTPSSPCGLYHQYHLAGHYA 180
QY 181 SATLWVAFCFWLLSNVLLSTPAPLYGGLALLTTGAFALFGVAFALASISSVPLCPLRLGSS 240
DB 181 SATLWVAFCFWLLSNVLLSTPAPLYGGLALLTTGAFALFGVAFALASISSVPLCPLRLGSS 240
QY 241 ALTTQGAAFWVTIATGVLCLFLGGAVVSLOQYVRPSALRTLLDQSAKDCSQERGSSPLIL 300
DB 241 ALTTQGAAFWVTIATGVLCLFLGGAVVSLOQYVRPSALRTLLDQSAKDCSQERGSSPLIL 300
QY 301 GDPLHKQAAALPDLKCITTNL 320
DB 301 GDPLHKQAAALPDLKCITTNL 320

RESULT 3

US-10-106-698-7554
; Sequence 7554, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Color and Colon Cancer Associated Polynucleotides and Polypepi
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 7554
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (65)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (85)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (88)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (98)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-7554

Query Match 5.9%; Score 19; DB 14; Length 98;
Best Local Similarity 100.0%; Pred. No. 7.1e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 RWFVLVRLVLSLFIGAIV 68
DB 8 RWFVLVRLVLSLFIGAIV 26

RESULT 4

US-09-759-130B-427
; Sequence 427, Application US/09759130B
; Publication No. US20030022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: Kirst, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodearl, Andrew
; APPLICANT: Holtzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: MPI00-535OMNIM
; CURRENT APPLICATION NUMBER: US/09/759,130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27

```

; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 427
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-130B-427
```

```

Query Match          5.9%; Score 19; DB 10; Length 112;
Best Local Similarity 100.0%; Pred. No. 8e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      145 ALEKGLPDPVLYLAEKFTP 163
      |||||
Db       75 ALEKGLPDPVLYLAEKFTP 93
```

```

RESULT 5
US-10-042-431-57
; Sequence 57, Application US/10042431
; Publication No. US20020182675A1
; GENERAL INFORMATION:
; APPLICANT: MCCARTHY, Sean A
; APPLICANT: BARNES, Thomas M
; APPLICANT: FRASER, Christopher C
; APPLICANT: SHARP, John D
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC,
; TITLE OF INVENTION: PREVENTIVE, THERAPEUTIC, AND OTHER USES
; FILE REFERENCE: 10147-6U2
; CURRENT APPLICATION NUMBER: US/10/042,431
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 57
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-042-431-57
```

```

Query Match          5.9%; Score 19; DB 13; Length 112;
Best Local Similarity 100.0%; Pred. No. 8e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      145 ALEKGLPDPVLYLAEKFTP 163
      |||||
Db       75 ALEKGLPDPVLYLAEKFTP 93
```

```

RESULT 6
US-10-094-749-2834
; Sequence 2834, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
```

```

; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2834
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2834
```

```

Query Match          5.9%; Score 19; DB 15; Length 298;
Best Local Similarity 100.0%; Pred. No. 2e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      145 ALEKGLPDPVLYLAEKFTP 163
      |||||
Db       100 ALEKGLPDPVLYLAEKFTP 118
```

```

RESULT 7
US-09-759-130B-425
; Sequence 425, Application US/09759130B
; Publication No. US20030022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: MCCARTHY, Sean A
; APPLICANT: FRASER, Christopher C
; APPLICANT: SHARP, John D
; APPLICANT: BARNES, Thomas S
; APPLICANT: KIRST, Susan J
; APPLICANT: MACKAY, Charles R
; APPLICANT: MYERS, Paul S
; APPLICANT: LEIBY, Kevin R
; APPLICANT: WRIGHTON, Nicolas
; APPLICANT: GOODEARL, Andrew
; APPLICANT: HOLTZMAN, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; TITLE OF INVENTION: USES.
; FILE REFERENCE: MPI00-5350NMIM
; CURRENT APPLICATION NUMBER: US/09/759,130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
```

```
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 425
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-759-130B-425
```

```
Query Match          5.9%; Score 19; DB 10; Length 343;
Best Local Similarity 100.0%; Pred. No. 2.3e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      145 ALEKGLDPVLYLAEKPTP 163
      |||||
DB      145 ALEKGLDPVLYLAEKPTP 163
```

```
RESULT 8
US-10-042-431-55
; Sequence 55, Application US/10042431
; Publication No. US20020182675A1
; GENERAL INFORMATION:
; APPLICANT: MCCARTHY, Sean A
; APPLICANT: BARNES, Thomas M
; APPLICANT: FRASER, Christopher C
; APPLICANT: SHARP, John D
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC,
; TITLE OF INVENTION: PREVENTIVE, THERAPEUTIC, AND OTHER USES
; FILE REFERENCE: 10147-6U2
; CURRENT APPLICATION NUMBER: US/10/042,431
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 55
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-042-431-55
```

```
Query Match          5.9%; Score 19; DB 13; Length 343;
Best Local Similarity 100.0%; Pred. No. 2.3e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      145 ALEKGLDPVLYLAEKPTP 163
      |||||
DB      145 ALEKGLDPVLYLAEKPTP 163
```

```
RESULT 9
US-09-764-891-2983
; Sequence 2983, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
```

```
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2983
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-764-891-2983
```

```
Query Match          2.5%; Score 8; DB 10; Length 49;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      55 VRVLSLF 62
      |||||
DB      39 VRVLSLF 46
```

```
RESULT 10
US-10-156-761-12862
; Sequence 12862, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12862
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-12862
```

```
Query Match          2.5%; Score 8; DB 14; Length 120;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      263 LGGAVVSL 270
      |||||
DB      34 LGGAVVSL 41
```

```
RESULT 11
US-10-128-714-8106
; Sequence 8106, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus an
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
```



```
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8106
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8106

Query Match
Best Local Similarity 100.0%; Score 8; DB 14; Length 345;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 VRVLSTLF 62
Db 46 VRVLSTLF 53

RESULT 12
US-10-109-791A-20
; Sequence 20, Application US/10109791A
; Publication No. US20030138787A1
; GENERAL INFORMATION:
; APPLICANT: BitTech Oncologic Corp.
; TITLE OF INVENTION: Functional Genetic Tests of DNA Mismatch Repair
; FILE REFERENCE: BTOL.102 NP
; CURRENT APPLICATION NUMBER: US/10/109,791A
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 315
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 615
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-109-791A-20

Query Match
Best Local Similarity 100.0%; Score 8; DB 14; Length 615;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 98 ALASISSV 105

RESULT 13
US-09-815-242-10961
; Sequence 10961, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10961
; LENGTH: 629
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-10961

Query Match
Best Local Similarity 100.0%; Score 8; DB 9; Length 629;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 98 ALASISSV 105

RESULT 14
US-09-815-242-12077
; Sequence 12077, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12077
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12077

Query Match
Best Local Similarity 100.0%; Score 8; DB 9; Length 633;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 223 ALASISSV 230
Db 102 ALASISSV 109

RESULT 15

US-09-815-242-5262
; Sequence 5262, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5262
; LENGTH: 669
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5262

Query Match 2.5%; Score 8; DB 9; Length 669;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ALASISSV 230
Db 99 ALASISSV 106

Search completed: February 23, 2004, 19:36:43
Job time : 37 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:20:36 : Search time 45 Seconds
(without alignments)
2243.685 Million cell updates/sec

Title: US-09-936-456-2
Perfect score: 320
Sequence: 1 MTLNMGVLPFPYQPRHAGF.....GDFLHKQALPDLKICITNTL 320

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	36	11.2	320	11	Q9D311	Q9d311 mus musculu
2	33	10.3	320	11	Q9D7U8	Q9d7u8 mus musculu
3	19	5.9	298	4	Q96M14	Q96m14 homo sapien
4	19	5.9	341	11	Q8VE49	Q8ve49 mus musculu
5	19	5.9	483	4	Q8N6K9	Q8n6k9 homo sapien
6	8	2.5	96	16	Q8XN05	Q8xn05 clostridium
7	8	2.5	120	16	Q82CL7	Q82cl7 streptomyce
8	8	2.5	141	16	Q827B8	Q827b8 salmonella
9	8	2.5	145	4	Q8N2Q6	Q8n2q6 homo sapien
10	8	2.5	227	16	Q914L7	Q914l7 pseudomonas
11	8	2.5	253	2	Q54145	Q54145 streptomyce
12	8	2.5	338	16	Q83P96	Q83p96 shigella fl
13	8	2.5	382	5	Q7YTN3	Q7ytn3 caenorhabdi
14	8	2.5	417	3	Q9UR09	Q9ur09 schizosacch
15	8	2.5	468	2	Q9L4P6	Q9l4p6 staphylococ
16	8	2.5	504	16	Q98E44	Q98e44 rhizobium 1

17	8	2.5	507	16	Q8DMD2	Q8dm2 synechococc
18	8	2.5	575	16	Q9I642	Q9i642 pseudomonas
19	8	2.5	598	16	Q7VN43	Q7vn43 haemophilus
20	8	2.5	615	16	Q7UAL0	Q7ual0 shigella fl
21	8	2.5	616	16	Q83P35	Q83p35 shigella fl
22	8	2.5	631	16	Q8EJ70	Q8ej70 shewanella
23	8	2.5	632	16	Q88DD1	Q88dd1 pseudomonas
24	8	2.5	645	16	Q8CPE9	Q8cpe9 staphylococ
25	8	2.5	645	16	Q87VJ2	Q87vj2 pseudomonas
26	8	2.5	725	2	Q52978	Q52978 rhizobium m
27	8	2.5	734	16	Q8NSL6	Q8nsl6 corynebacte
28	8	2.5	850	16	Q89NM3	Q89nm3 bradyrhizob
29	8	2.5	926	2	Q8RTB9	Q8rte9 rhizobium m
30	8	2.5	999	16	Q92LV7	Q92lv7 rhizobium m
31	7	2.2	15	6	Q28822	Q28822 oryctolagus
32	7	2.2	29	4	Q8TAI5	Q8tai5 homo sapien
33	7	2.2	30	6	Q8MHM1	Q8mhw1 sus scrofa
34	7	2.2	56	6	Q8MHX9	Q8mhx9 sus scrofa
35	7	2.2	58	2	Q93NM4	Q93nw4 streptococc
36	7	2.2	72	16	Q7U7C7	Q7u7c7 synechococc
37	7	2.2	86	2	Q9EZ16	Q9ez16 sodalis glo
38	7	2.2	86	2	Q93RC2	Q93rc2 escherichia
39	7	2.2	86	2	Q8GAT3	Q8gat3 primary end
40	7	2.2	86	16	Q8X6F2	Q8xf2 escherichia
41	7	2.2	93	16	Q92Z62	Q92z62 rhizobium m
42	7	2.2	97	2	Q9EY73	Q9eyt3 pseudomonas
43	7	2.2	100	16	Q99YG0	Q99yg0 streptococc
44	7	2.2	100	16	Q8NZU6	Q8nzu6 streptococc
45	7	2.2	102	17	Q9YCC34	Q9yc34 aeropyrum p.

ALIGNMENTS

RESULT 1

ID	Q9D311	PRELIMINARY;	PRT;	320 AA.
AC	Q9D311;			
DT	01-JUN-2001 (TREMBlrel. 17, Created)			
DT	01-JUN-2001 (TREMBlrel. 17, Last sequence update)			
DT	01-OCT-2002 (TREMBlrel. 22, Last annotation update)			
DE	9030623N16R1k protein (RIKEN cDNA 9030623N16 gene).			
GN	9030623N16R1K.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
CX	NCBI_Taxid=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Colon;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Guernicich S., Hill D., Hofman M., Hume D.A., Kamiya M., Lee N.H.,			
RA	Lyon P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyokawa K., Wang K.H., Weltz C., Whitaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,			
RA	Hayashizaki Y.;			
RT	"Functional annotation of a full-length mouse cDNA collection.";			
RL	Nature 409:685-690(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			

RC TISSUE=Colon;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK018569; BAB31281.1; -
DR EMBL; BC031111; AAH31111.1; -
DR MGD; MGI:1914061; 9030623N16R1K.
SQ SEQUENCE 320 AA; 35427 MW; 4B7D4C81685162F7 CRC64;

Query Match 11.2%; Score 36; DB 11; Length 320;
Best Local Similarity 100.0%; Pred. No. 6.7e-26;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 ALEKGLPDPVLYLAKEFTPSPGGLYHQYHLAGHYA 180
Db 145 ALEKGLPDPVLYLAKEFTPSPGGLYHQYHLAGHYA 180

RESULT 2

ID Q9D7U8 PRELIMINARY; PRT; 320 AA.

AC Q9D7U8;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE 9030623N16R1K protein.

GN 9030623N16R1K.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Stomach;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., Kling B., Kochwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,

RA Schriml L.M., Staahl F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Guestincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Marzarelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,

RA Wyrnshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).

DR EMBL; AK008816; BAB25910.1; -

DR MGD; MGI:1914061; 9030623N16R1K.

SQ SEQUENCE 320 AA; 36342 MW; 9541848376D0E2E2 CRC64;

Query Match 10.3%; Score 33; DB 11; Length 320;
Best Local Similarity 100.0%; Pred. No. 5.2e-23;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 ILPGIRGSRWFWLVRLSLFAGAEIVAVHFS 73
Db 41 ILPGIRGSRWFWLVRLSLFAGAEIVAVHFS 73

RESULT 3

ID Q96M14

AC Q96M14 PRELIMINARY; PRT; 298 AA.

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)

DE Hypothetical protein FLJ32334.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Prostate;

RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,

RA Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J., Chiba Y.,

RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,

RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,

RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,

RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,

RA Wagsatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,

RA Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isogai T.,

RT "NEDO human cDNA sequencing project.";

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Lung;

RA Strausberg R.;

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AK056896; BAB71304.1; -

DR EMBL; BC020841; AAH20841.1; -

KW Hypothetical protein.

SQ SEQUENCE 298 AA; 33110 MW; D9B346C25D142E1D CRC64;

Query Match 5.9%; Score 19; DB 4; Length 298;
Best Local Similarity 100.0%; Pred. No. 1.4e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 ALEKGLPDPVLYLAKEFTP 163
Db 100 ALEKGLPDPVLYLAKEFTP 118

RESULT 4

ID Q8VE49

AC Q8VE49 PRELIMINARY; PRT; 341 AA.

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Hypothetical protein.

GN BC019755.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Strausberg R.;

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC019755; AAH19755.1; -

DR MGD; MGI:2384861; BC019755.

KW Hypothetical protein.

SQ SEQUENCE 341 AA; 37595 MW; CC16D7AD103A7276 CRC64;

Query Match 5.9%; Score 19; DB 11; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 ALEKGLPDPVLYLAKEFTP 163
Db 145 ALEKGLPDPVLYLAKEFTP 163

RESULT 5

ID Q8N6K9

AC Q8N6K9 PRELIMINARY; PRT; 483 AA.

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DE Hypothetical protein.

GN BC019755.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Strausberg R.;

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC019755; AAH19755.1; -

DR MGD; MGI:2384861; BC019755.

KW Hypothetical protein.

SQ SEQUENCE 341 AA; 37595 MW; CC16D7AD103A7276 CRC64;

```

DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Similar to RIKEN CDNA 9030623N16 gene.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Plrimates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Lung;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC029819; AAH29819.1; -
SQ SEQUENCE 483 AA; 53519 MW; 54BEF64528B96C18 CRC64;

Query Match
Best Local Similarity 5.9%; Score 19; DB 4; Length 483;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 ALEKGLPDPVLYLAEKFTP 163
Db 145 ALEKGLPDPVLYLAEKFTP 163

RESULT 6
Q8XN05 PRELIMINARY; PRT; 96 AA.
ID Q8XN05;
AC Q8XN05;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein CPE0277.
GN CPE0277.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003186; BAB79983.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 96 AA; 10261 MW; 4FF6184BA188D076 CRC64;

Query Match
Best Local Similarity 2.5%; Score 8; DB 16; Length 96;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 LAASFLLI 41
Db 9 LAASFLLI 16

RESULT 7
Q82CL7 PRELIMINARY; PRT; 120 AA.
ID Q82CL7;
AC Q82CL7;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein.
GN SAV5328.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005042; BAC73040.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 120 AA; 13549 MW; CEB73465B174F973 CRC64;

Query Match
Best Local Similarity 2.5%; Score 8; DB 16; Length 120;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 LGAVVSL 270
Db 34 LGAVVSL 41

RESULT 8
Q827B8 PRELIMINARY; PRT; 141 AA.
ID Q827B8;
AC Q827B8;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical periplasmic protein STY1364.
GN STY1364 OR T1602.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltham T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Pary C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Lion S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
DR EMBL; AL627270; CAD01632.1; -
DR EMBL; AF016839; AA069231.1; -
KW Hypothetical protein; Complete proteome.

```

SEQ SEQUENCE 141 AA; 15652 MW; BA6565B29A3B8844 CRC64;
Query Match 2.5%; Score 8; DB 16; Length 141;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 223 ALA5ISSV 230
DB 11 ALA5ISSV 18
RESULT 9
Q8N2Q6 PRELIMINARY; PRT; 145 AA.
AC Q8N2Q6;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ90047.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuma S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT "NEDO human cDNA sequencing project."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK074528; BAC11041.1; -.
KM Hypothetical protein.
SQ SEQUENCE 145 AA; 16153 MW; EF44C8E658ECAC2E CRC64;
Query Match 2.5%; Score 8; DB 4; Length 145;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 197 L1STPAPL 204
DB 42 L1STPAPL 49
RESULT 10
Q914L7 PRELIMINARY; PRT; 227 AA.
AC Q914L7;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Hypothetical protein PA1118.
GN PA1118.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
opportunistic pathogen."
RL Nature 406:959-964 (2000).
DR EMBL; AE004542; AAG04507.1; -.

DR PIR; B83505; B83505.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 227 AA; 25612 MW; FAD67BC0A50B9EC CRC64;
Query Match 2.5%; Score 8; DB 16; Length 227;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 98 RVGLVGL 105
DB 105 RVGLVGL 112
RESULT 11
Q54145 PRELIMINARY; PRT; 253 AA.
AC Q54145;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Thioesterase.
OS Streptomyces fradiae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1906;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TS9235;
RX MEDLINE=95075319; PubMed=7984112;
RA Merson-Davies L.A., Cundliffe E.;
RT "Analysis of five tylosin biosynthetic genes from the tyl1BA region of
the Streptomyces fradiae genome."
RL Mol. Microbiol. 13:349-355 (1994).
DR EMBL; U08223; AAA21345.1; -.
DR PIR; S49055; S49055.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0016788; F:hydrolase activity, acting on ester bonds; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR000379; Ser esters.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00975; Thioesterase; 1.
SQ SEQUENCE 253 AA; 27684 MW; 8AB752B2CDB1450F CRC64;
Query Match 2.5%; Score 8; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 211 LTTGAFAL 218
DB 213 LTTGAFAL 220
RESULT 12
Q83P96 PRELIMINARY; PRT; 338 AA.
AC Q83P96;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Periplasmic binding protein component of pn transporter (Periplasmic
binding protein component of phosphonate transporter).
GN PHND OR SF4118 OR S3612.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,

RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity
RT through comparison with genomes of *Escherichia coli* K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of *Shigella*
RT *flexneri* serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL; AB015420; AAN45543.1; -.
DR EMBL; AB016989; AAP18656.1; -.
DR GO; GO:0030288; C:periplasmic space (sensu Gram-negative Bact. .; IEA.
DR GO; GO:0015604; F:phosphonate transporter activity; IEA.
DR GO; GO:0015716; P:phosphonate transport; IEA.
DR InterPro; IPR005770; Phosphonate bind.
DR TIGRfams; TIGR01098; 3A0109s03R; 1.
KM Complete proteome.
SQ SEQUENCE 338 AA; 37382 MW; 9A9C3C3861CBE594 CRC64;

QY Query Match 2.5%; Score 8; DB 16; Length 338;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 225 SPLITGDP 232
296 SPLITGDP 303
|||||
|

RESULT 13
Q7YTN3 PRELIMINARY; PRT; 382 AA.
ID Q7YTN3;
AC Q7YTN3;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE C. elegans SRW-42 protein (Corresponding sequence F14F8.11).
GN SRW-42.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Lloyd C.R.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z92782; CAB17813.1; -.
SQ SEQUENCE 382 AA; 43552 MW; 492BCAA4F37C4D1C CRC64;

QY Query Match 2.5%; Score 8; DB 5; Length 382;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 56 RVLSLFI 63
|||||
| 97 RVLSLFI 104

RESULT 14
Q9UR09 PRELIMINARY; PRT; 417 AA.
ID Q9UR09;
AC Q9UR09;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Cell wall synthesis protein PSU1.
GN PSU1 OR PSU1+.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Cadieu E., Lelaure V., Galibert F., McDougall R.C., Rajandream M.A.,
RA Barrell B.G.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RX MEDLINE=99393312; PubMed=10462482;
RA Omi K., Sonoda H., Nagata K., Sugita K.;
RT "Cloning and characterization of psu1+, a new essential fission yeast
RT gene involved in cell wall synthesis.";
RL Biochem. Biophys. Res. Commun. 262:368-374(1999).
DR EMBL; AL136078; CAB65613.1; -.
DR EMBL; AB009980; BAA83907.1; -.
DR PIR; JC7092; JC7092.
DR InterPro; IPR005556; SUN.
DR Pfam; PF03856; SUN; 1.
SQ SEQUENCE 417 AA; 41750 MW; 72BCAEE2226208CB CRC64;

QY Query Match 2.5%; Score 8; DB 3; Length 417;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 208 LALLTGA 215
|||||
| 9 LALLTGA 16
|

RESULT 15
Q9LAP6 PRELIMINARY; PRT; 468 AA.
ID Q9LAP6;
AC Q9LAP6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE DNA mismatch repair protein (Fragment).
GN MUTL.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COL;
RX MEDLINE=20031141; PubMed=10566865;
RA de Lencastre H., Wu S.W., Pinho M.G., Ludovice A.M., Filipe S.,
RA Gardete S., Sobral R., Gill S., Chung M., Tomasz A.;
RT "Antibiotic resistance as a stress response: complete sequencing of a
RT large number of chromosomal loci in *Staphylococcus aureus* strain COL
RT that impact on the expression of resistance to methicillin.";
RL Microb. Drug Resist. 5:163-175(1999).
DR EMBL; AJ131755; CAB82461.1; -.
DR HSSP; P23367; 1BKN.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0006298; P:mismatch repair; IEA.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR002099; DNA_mis_repair.
DR Pfam; PF01119; DNA_mis_repair; 1.

DR Pfam; PF02518; HATPase_c; 1.
DR SMART; SM00387; HATPase_c; 1.
DR TIGRPFAMs; TIGR00585; multi; 1.
DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
FT NON_TER 468 468
SQ SEQUENCE 468 AA; 52987 MW; 87BADB2D50109FBE CRC64;

Query Match 2.5%; Score 8; DB 2; Length 468;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 223 ALASISSV 230
Db 99 ALASISSV 106

Search completed: February 23, 2004, 19:34:47
Job time : 47 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 18:16:40 ; Search time 98 Seconds

(without alignments)
922.604 Million cell updates/sec

Title: US-09-936-456-2

Perfect score: 1666
Sequence: 1 MTLMNGVLPFYDQPRHAGF.....GDP LHKQALPDLKCTTNL 320

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1666	100.0	320	3 AAB18992	Aab18992 Amino aci
2	1666	100.0	320	3 AAB26325	Aab26325 Human CAS
3	1660	99.6	320	5 ABB90126	Abb90126 Human pol
4	1660	99.6	370	7 ADD19303	Add19303 Human sec
5	1253	75.2	242	3 AAB52095	Aab52095 Human sec
6	908	54.5	262	3 AAB26326	Aab26326 CASB618 p
7	883	53.0	343	5 ABB97510	Abb97510 Novel hum
8	874	52.5	343	4 AAB66071	Aab66071 Human INT
9	874	52.5	343	6 ABU08370	Abu08370 Amino aci
10	874	52.5	343	6 ABO32725	Abob32725 Secreted
11	697.5	41.9	298	6 ADA55266	Ada55266 Human pro
12	460.5	27.6	445	4 ABG03520	Abg03520 Novel hum
13	458	27.5	112	4 AAB66073	Aab66073 Human INT
14	458	27.5	112	6 ABO32727	Abob32727 Secreted
15	276.5	16.6	383	4 AAB66060	Aab66060 C06E1.3 r
16	276.5	16.6	383	6 ABO32714	Abob32714 Secreted
17	254.5	15.3	363	4 ABB60704	Abb60704 Drosophil
18	254.5	15.3	363	4 ABB67423	Abb67423 Drosophil
19	254	15.2	103	7 ADD19262	Add19262 Human sec
20	242.5	14.6	253	3 AAB52094	Aab52094 Gene 44 h
21	110	6.6	450	6 ABU21931	Abu21931 Protein e
22	109	6.5	369	4 AAB20381	Abab20381 Soybean s
23	108	6.5	648	6 ABU32147	Abu32147 Protein e
24	107	6.4	671	6 ABU28385	Abu28385 Protein e
25	106.5	6.4	374	4 AAB20379	Aab20379 Corn ster

26	105.5	6.3	473	2 AAY38688	Aay38688 Neisseria
27	105.5	6.3	473	3 AAY75470	Aay75470 Neisseria
28	105.5	6.3	473	6 ABP77672	Abp77672 N. gonorr
29	104	6.2	364	5 ABG97436	Abg97436 S. carzin
30	104	6.2	652	4 AAU38223	Aau38223 Salmonell
31	104	6.2	652	6 AAU47915	Aau47915 Protein e
32	103.5	6.2	428	6 ABU21684	Abu21684 Protein e
33	103	6.2	98	4 AAG76780	Aag76780 Human col
34	101.5	6.1	428	6 ABM67285	Abm67285 Photorhab
35	101.5	6.1	465	6 ABU48052	Abu48052 Protein e
36	100.5	6.0	431	4 ABB59558	Abb59558 Drosophil
37	100.5	6.0	585	2 AAY38798	Aay38798 Neisseria
38	99.5	6.0	477	4 AAU38247	Aau38247 Salmonell
39	99.5	6.0	533	6 ADA36605	Ada36605 Acinetoba
40	99.5	6.0	558	2 AAY38799	Aay38799 Neisseria
41	99	5.9	612	6 ABU31012	Abu31012 Protein e
42	98.5	5.9	317	3 AAY57097	Aay57097 UDP-glucu
43	98.5	5.9	464	4 ABB52659	Abb52659 Escherich
44	98.5	5.9	481	6 ABU45403	Abu45403 Protein e
45	98	5.9	463	4 AAU34551	Aau34551 E. coli c

ALIGNMENTS

RESULT 1	
ID	AAB18992 standard; protein; 320 AA.
AC	AAB18992;
DT	08-FEB-2001 (first entry)
DE	Amino acid sequence of a human transmembrane protein.
XX	Human; transmembrane protein; cell proliferation disorder; myeloma;
XX	reproductive disorder; smooth muscle disorder; neurological disorder;
KW	arteriosclerosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
KW	allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;
KW	Alzheimer's disease; Tourette's disorder.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Modified-site
FT	Location/Qualifiers
FT	84
FT	/note= "potential glycosylation site"
FT	86
FT	/note= "potential phosphorylation site"
FT	96
FT	/note= "potential phosphorylation site"
FT	109
FT	/note= "potential glycosylation site"
FT	121
FT	/note= "potential glycosylation site"
FT	131
FT	/note= "potential glycosylation site"
FT	136
FT	/note= "potential phosphorylation site"
FT	156
FT	/note= "potential glycosylation site"
FT	280
FT	/note= "potential phosphorylation site"
FT	285
FT	/note= "potential phosphorylation site"
XX	
PN	WO200056891-A2.
XX	
PD	28-SEP-2000.
XX	
PF	22-MAR-2000; 2000MO-US007817.
XX	
PR	22-MAR-1999; 99US-0125537P.
PR	16-JUN-1999; 99US-0139565P.
XX	
PA	(INCY-) INCYTE PHARM INC.

XX Yue H, Lal P, Tang YT, Hillman JL, Reddy R, Bandman O;
PI Baughn MR, Lu DAM, Azimzai Y, Yang J;
XX WPI; 2000-579485/54.
DR N-PSDB; AAA96505.
XX
PT New human transmembrane proteins are used to treat a disease or condition
PT associated with decreased expression of functional HTMP e.g. Tourette's
PT disorder, angina and leukemia.
XX
PS Claim 1; Page 109; 130pp; English.
XX
CC The present sequence represents a human transmembrane proteins (HTMP).
CC Agonists and antagonists of the protein are used to treat a disease or
CC condition associated with overexpression of the protein. Diseases and
CC conditions which can be treated include include cell proliferative,
CC immunological, reproductive, smooth muscle and neurological disorders
CC e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency
CC syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,
CC stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The
CC polynucleotides may be used to detect and quantify gene expression in
CC biopsied tissues where protein expression may be correlated with disease
CC e.g. to determine absence, presence or excess expression of HTMP or to
CC monitor regulation of HTMP expression during therapeutic intervention
XX
SQ Sequence 320 AA;

Query Match 100.0%; Score 1666; DB 3; Length 320;
Best Local Similarity 100.0%; Pred. No. 2.1e-173;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTLWNGVLPFYPPQPRHAAGFSVPLILVILVFLALAAFLILPGIRGSRMFWLVRLLS 60
DB 1 MTLWNGVLPFYPPQPRHAAGFSVPLILVILVFLALAAFLILPGIRGSRMFWLVRLLS 60
QY 61 LFIGAEIVAVHFSAEWFGVTNTNTSYKAFSARVTARVGLVGLGGINITLTGTPVHQL 120
DB 61 LFIGAEIVAVHFSAEWFGVTNTNTSYKAFSARVTARVGLVGLGGINITLTGTPVHQL 120
QY 121 NETIDYNEQFTWRLEKENVAAEYANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYA 180
DB 121 NETIDYNEQFTWRLEKENVAAEYANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYA 180
QY 181 SATLWVAFCFWLISNVLLSTPAPLYGGLALLTTGAFALFGVAFALASISSVPLCPRLGSS 240
DB 181 SATLWVAFCFWLISNVLLSTPAPLYGGLALLTTGAFALFGVAFALASISSVPLCPRLGSS 240
QY 241 ALTYQGAFFWTLATGVLCLFLGGAVVSLOQYRPSALRTLDDQSAKDCSQERGSSPLIL 300
DB 241 ALTYQGAFFWTLATGVLCLFLGGAVVSLOQYRPSALRTLDDQSAKDCSQERGSSPLIL 300
QY 301 GDPILHQAALLPDLKICITTNL 320
DB 301 GDPILHQAALLPDLKICITTNL 320

RESULT 2
AAB26325
ID AAB26325 standard; protein; 320 AA.
XX
AC AAB26325;
XX
DT 11-JAN-2001 (first entry)
XX
DE Human CASB618 protein.
XX
KW Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian;
KW colon; autoimmune disease; HLA_A0201.
XX
OS Homo sapiens.
XX
PN WO200053748-A2.

XX PD 14-SEP-2000.
XX
XX PF 09-MAR-2000; 2000WO-EP002048.
XX
XX PR 11-MAR-1999; 99GB-00005607.
XX PR 01-SEP-1999; 99GB-00020590.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX PI Bruck CEM, Cassart J, Coche T, Vinals Y De Baseolsc;
XX
XX WPI; 2000-572268/53.
DR N-PSDB; AAA94623.
XX
PT New human CASB618 polypeptide, useful as a vaccine for prophylactic and
PT therapeutic treatment of cancers, particularly ovarian or colon cancer,
PT autoimmune diseases and related conditions.
XX
PS Claim 1; Page 61-62; 76pp; English.
XX
CC The present sequence is human CASB618 protein. The gene for human CASB618
CC is thought to be located on chromosome 15. The present protein and
CC epitopes of this protein (see AAB26327 to AAB26399) are useful in
CC diagnosing the occurrence of tumour cells and in vaccines for prophylactic
CC and therapeutic treatment of cancers, particularly ovarian or colon
CC cancer, autoimmune diseases and related conditions
XX
SQ Sequence 320 AA;

Query Match 100.0%; Score 1666; DB 3; Length 320;
Best Local Similarity 100.0%; Pred. No. 2.1e-173;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTLWNGVLPFYPPQPRHAAGFSVPLILVILVFLALAAFLILPGIRGSRMFWLVRLLS 60
DB 1 MTLWNGVLPFYPPQPRHAAGFSVPLILVILVFLALAAFLILPGIRGSRMFWLVRLLS 60
QY 61 LFIGAEIVAVHFSAEWFGVTNTNTSYKAFSARVTARVGLVGLGGINITLTGTPVHQL 120
DB 61 LFIGAEIVAVHFSAEWFGVTNTNTSYKAFSARVTARVGLVGLGGINITLTGTPVHQL 120
QY 121 NETIDYNEQFTWRLEKENVAAEYANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYA 180
DB 121 NETIDYNEQFTWRLEKENVAAEYANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYA 180
QY 181 SATLWVAFCFWLISNVLLSTPAPLYGGLALLTTGAFALFGVAFALASISSVPLCPRLGSS 240
DB 181 SATLWVAFCFWLISNVLLSTPAPLYGGLALLTTGAFALFGVAFALASISSVPLCPRLGSS 240
QY 241 ALTYQGAFFWTLATGVLCLFLGGAVVSLOQYRPSALRTLDDQSAKDCSQERGSSPLIL 300
DB 241 ALTYQGAFFWTLATGVLCLFLGGAVVSLOQYRPSALRTLDDQSAKDCSQERGSSPLIL 300
QY 301 GDPILHQAALLPDLKICITTNL 320
DB 301 GDPILHQAALLPDLKICITTNL 320

RESULT 3
ABB90126
ID ABB90126 standard; protein; 320 AA.
XX
AC ABB90126;
XX
DT 24-MAY-2002 (first entry)
XX
DE Human polypeptide SEQ ID NO 2502.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;

KM neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX WO200190304-A2.
XX
XX 29-NOV-2001.
XX
XX 18-MAY-2001; 2001WO-US016450.
XX
XX 19-MAY-2000; 2000US-0205515P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Birse CE, Rosen CA;
XX
XX WPI; 2002-122018/16.
XX
XX N-PSDB; ABL90535.
XX
XX
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders.
XX
XX
XX Claim 11; SEQ ID NO 2502; 2081pp + Sequence listing; English.
XX
XX
XX The invention relates to novel genes (AB189449-AB190853) and proteins
CC (AB889040-AB890444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX
SQ Sequence 320 AA;
XX
XX
XX Query Match 99.6%; Score 1660; DB 5; Length 320;
XX Best Local Similarity 99.7%; Pred. No. 9.6e-173;
XX Matches 319; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 MTLMNGVLPFYPPQPHHAGFVSVLLIIVLFALAAFFLLIPGIRGSHRMFLVRVLIS 60
Db 1 MTLMNGVLPFYPPQPHHAGFVSVLLIIVLFALAAFFLLIPGIRGSHRMFLVRVLIS 60
QY 61 LFIGAEIVAVHFSAEWFGVTNNTSYKAFSARVTARVGLLVGLEGINITLTGTPVHQL 120
Db 61 LFIGAEIVAVHFSAEWFGVTNNTSYKAFSARVTARVGLLVGLXGINITLTGTPVHQL 120
QY 121 NETIDVNEQFTWRLEKYNAAEYANALEKGLPDPVLYLAEKFTPSSPCGLHYOYHLA GHA 180
Db 121 NETIDVNEQFTWRLEKYNAAEYANALEKGLPDPVLYLAEKFTPSSPCGLHYOYHLA GHA 180
QY 181 SATLWVAFCFWLLSNVLLSTPAPLYGGIALITTGAFALFGVAFALASISSVPLCPRLGSS 240
Db 181 SATLWVAFCFWLLSNVLLSTPAPLYGGIALITTGAFALFGVAFALASISSVPLCPRLGSS 240
QY 241 ALTTQYGAAFWVTLTATGVLCLFLGGA VVSLQYVRPSALRTLLDOSAKDCSQRGGSPIL 300
Db 241 ALTTQYGAAFWVTLTATGVLCLFLGGA VVSLQYVRPSALRTLLDOSAKDCSQRGGSPIL 300
QY 301 GDPLHKQAAALPDLKCTITNTL 320
Db 301 GDPLHKQAAALPDLKCTITNTL 320

RESULT 4
ADD19303
ID ADD19303 standard; protein; 370 AA.
XX
XX ADD19303;
AC
XX
DT 15-JAN-2004 (first entry)
XX
DE Human secreted protein from gene 4 #2.
XX
KM human secreted protein; cyrostatic; antibacterial; virucide;
KM neuroprotective; gynaecological; gastrointestinal-Gen; cardiant;
KM cardiovascular-Gen; nephrotropic; antiinflammatory; muscular-Gen;
KM respiratory-Gen; immunosuppressive; cerebroprotective; vasotropic;
KM nootropic; antiallergic; cancer; bacterial infection; viral infection;
KM neural disorder; immune system disorder; blood disorder;
KM muscular disorder; reproductive disorder; gastrointestinal disorder;
KM pulmonary disorder; cardiovascular disorder; renal disorder;
KM inflammatory disorder; proliferative disorder; human.
XX
XX
OS Homo sapiens.
XX
XX WO2003052377-A2.
XX
XX 26-JUN-2003.
XX
XX 06-NOV-2002; 2002WO-US035606.
XX
XX 07-NOV-2001; 2001US-0331046P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2003-533050/50.
XX
XX N-PSDB; ADD19228.
XX
XX
XX New isolated nucleic acids encoding signal transduction pathway component
PT polypeptides, useful for diagnosing, treating, and/or preventing
PT disorders, such as cancer, infections, cardiovascular and inflammatory
PT diseases.
XX
XX
XX Claim 11; SEQ ID NO 130; 554pp; English.
XX
XX
XX The invention relates to an isolated nucleic acid molecule (cDNA)
CC encoding a human secreted protein, representing one of 85 novel genes.
CC Also included are recombinant vectors, host cells (expressing the
CC protein), the secreted proteins (including their fragments, epitopes and
CC homologues), an isolated antibody that binds specifically to the protein,
CC diagnosing a pathological condition or susceptibility to a pathological
CC condition (comprising determining the presence or absence of a mutation
CC in the nucleic acid and diagnosing a condition based on the presence or
CC absence of the mutation), diagnosing a pathological condition or
CC susceptibility to a pathological condition (comprising determining the
CC presence or amount of expression of the protein in a biological sample
CC and diagnosing a condition based on the presence or amount of expression
CC of the protein), preventing, treating or ameliorating a medical condition
CC by administering the nucleic acid or protein to a mammalian subject,
CC identifying a binding partner to the protein, the gene corresponding to
CC (comprising expressing the nucleic acid in a cell, isolating the
CC supernatant, detecting an activity in a biological assay and identifying
CC the protein in the supernatant having the activity). The nucleic acids
CC and proteins display the following activities Cytostatic, antibacterial,
CC Virucide, Neuroprotective, Gynaecological, Gastrointestinal-Gen,
CC Cardiant, Cardiovascular-Gen, Nephrotropic, Antiinflammatory, Muscular-
CC Gen, Respiratory-Gen, Immunosuppressive, Cerebroprotective, Vasotropic,
CC Nootropic, Antiallergic. The methods and compositions of the present
CC invention are useful for diagnosing, treating, preventing and/or
CC prognosticating disorders related to the novel polypeptides, such as
CC cancer, bacterial or viral infections, and neural, immune system, blood,

CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal, inflammatory or proliferative disorders (many examples of these
CC diseases and disorders are given in the specification). The present
CC sequence represents a novel secreted protein of the invention.
XX
SQ Sequence 370 AA;

Query Match 99.6%; Score 1660; DB 7; Length 370;
Best Local Similarity 99.7%; Pred. No. 1.2e-172;
Matches 319; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTLWNGVLPFYPPQPRHAGFSVPLILVLFVFLAASFLILPGIRGSRWFVLRVLIS 60
DB 51 MTLWNGVLPFYPPQPRHAGFSVPLILVLFVFLAASFLILPGIRGSRWFVLRVLIS 110
QY 61 LFIGAEIVAVHFSAEWFGVTNTNTSYKAFSAARVTARVGLVLEGINITLTGTPVHQL 120
DB 111 LFIGAEIVAVHFSAEWFGVTNTNTSYKAFSAARVTARVGLVLEGINITLTGTPVHQL 170
QY 121 NETIDYNEQFTWRLKENYAAEYANALEKGLPDPVLYLAKEFTPSSPCGLYHGYHLAHHYA 180
DB 171 NETIDYNEQFTWRLKENYAAEYANALEKGLPDPVLYLAKEFTPSSPCGLYHGYHLAHHYA 230
QY 181 SATLWVAFCFWLLSNVLLSTPAPLYGGLALLTGAFALFGVAFALASISSVPLCPLRLGSS 240
DB 231 SATLWVAFCFWLLSNVLLSTPAPLYGGLALLTGAFALFGVAFALASISSVPLCPLRLGSS 290
QY 241 ALTYQGAAFWTLATGVLCFLGGAVVSLOQYVRPSALRTLDDQSAKDCQERGGSPILL 300
DB 291 ALTYQGAAFWTLATGVLCFLGGAVVSLOQYVRPSALRTLDDQSAKDCQERGGSPILL 350
QY 301 GDPFHQAALPDLKCTTTNL 320
DB 351 GDPFHQAALPDLKCTTTNL 370

RESULT 5
AAB52095
ID AAB52095 standard; protein; 242 AA.

AC AAB52095;
XX
DT 21-FEB-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 44 SEQ ID NO:144.
XX
KW Human; secreted protein; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antitumor; anticonvulsant; antibacterial;
KW antifungal; antiparasitic; cardiac; cancer; immune disease; allergy;
KW cardiovascular disorder; wound healing; infection; neurological disease.
XX
OS Homo sapiens.
XX
PN WO200061596-A1.
XX
PD 19-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US008983.
XX
PR 09-APR-1999; 99US-0128703P.
XX
PR 14-JAN-2000; 2000US-0176068P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Ruben SM, Komatsoulis G;
XX
DR WPI, 2000-611865/58.
DR N-PSDB; AAC95564.
XX
PT Fifty nucleic acid molecules encoding human secreted proteins, useful in
PT the prevention, treatment and diagnosis of cancer, immune disorders,

PT cardiovascular disorders and neurological diseases.
XX
PS Disclosure; Page 75-76; 505pp; English.
XX

CC Polynucleotide sequences AAC95521 - AAC95570 represent cDNA encoding 50
CC human secreted proteins AAB52012 - AAB52161. Sequences AAB52062 -
CC AAB52103 represent alternative polypeptides encoded by the genes, and
CC amino acid sequences with which they share homology. The genes and
CC proteins have activities dependent on the tissues and cells in which they
CC are expressed. Examples of their activities include cytostatic;
CC immunosuppressive; nootropic; neuroprotective; antiviral; antiallergic;
CC hepatotropic; antidiabetic; antiinflammatory; antitumor; vulnary;
CC anticonvulsant; antibacterial; antifungal; antiparasitic; and cardiac.
CC The secreted proteins, polypeptides, antagonists and agonists may be
CC useful in treating, preventing and/or diagnosing diseases and disorders
CC such as cancer, particularly breast and ovarian cancer, and other cancers
CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,
CC liver, lung, or urogenital. Immune disorders such as Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemia; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases such as viral,
CC bacterial, fungal and parasitic infections may also be treated using the
CC proteins and polynucleotides of the invention. Sequences AAC95512 -
CC AAC95520 and AAB52011 are used in the isolation and characterisation of
CC the proteins and polynucleotides of the invention

XX
SQ Sequence 242 AA;

Query Match 75.2%; Score 1253; DB 3; Length 242;
Best Local Similarity 99.6%; Pred. No. 2.4e-128;
Matches 241; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 44 GIRGHSRWFVLRVLLSLFIGAEIVAVHFSAEWFGVTNTNTSYKAFSAARVTARVGLV 103
DB 1 GIRGHSRWFVLRVLLSLFIGAEIVAVHFSAEWFGVTNTNTSYKAFSAARVTARVGLV 60
QY 104 GLEGINITLTGTPVHQNETIDYNEQFTWRLKENYAAEYANALEKGLPDPVLYLAKEFTP 163
DB 61 GLEGINITLTGTPVHQNETIDYNEQFTWRLKENYAAEYANALEKGLPDPVLYLAKEFTP 120
QY 164 SSPCGLYHGYHLAHHYASATLWVAFCFWLLSNVLLSTPAPLYGGLALLTGAFALFGVFA 223
DB 121 SSPCGLYHGYHLAHHYASATLWVAFCFWLLSNVLLSTPAPLYGGLALLTGAFALFGVFA 180
QY 224 LASISSVPLCPLRLGSSALLTYQGAAFWTLATGVLCFLGGAVVSLOQYVRPSALRTLDD 283
DB 181 LASISSVPLCPLRLGSSALLTYQGAAFWTLATGVLCFLGGAVVSLOQYVRPSALRTLDD 240
QY 284 QS 285
DB 241 QS 242

RESULT 6
AAB26326
ID AAB26326 standard; protein; 262 AA.

AC AAB26326;
XX
DT 12-SEP-2003 (revised)
DT 11-JAN-2001 (first entry)
XX
DE CASB618 protein.
XX
KW Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian;
KW colon; autoimmune disease; HLA_A0201.
XX
OS Homo sapiens.
OS Influenza virus.
OS Chimeric.
XX

PN WO200053748-A2.
 XX 14-SEP-2000.
 PD
 XX 09-MAR-2000; 2000WO-EP002048.
 PF
 XX 11-MAR-1999; 99GB-00005607.
 PR 01-SEP-1999; 99GB-00020590.
 XX
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PI Bruck CEM, Cassart J, Coche T, Vinals Y De Bassolac;
 XX
 DR WPI; 2000-572268/53.
 XX
 PT New human CASB618 polypeptide, useful as a vaccine for prophylactic and
 PT therapeutic treatment of cancers, particularly ovarian or colon cancer,
 PT autoimmune diseases and related conditions.
 XX
 PS Example 7; Page 63; 76pp; English.
 XX
 CC The present invention relates to epitopes of human CASB618 protein (see
 CC AAB26327 to AAB26399). The epitopes of AAB26327 to AAB26399 are useful in
 CC diagnosing the occurrence of tumour cells and in vaccines for prophylactic
 CC and therapeutic treatment of cancers, particularly ovarian or colon
 CC cancer, autoimmune diseases and related conditions. The present sequence
 CC is a protein derived from a chimeric gene constructed from the CASB618
 CC addition of an IFP (NS1 DNA sequence encoding the N-terminal 1-81 amino
 CC acids of the NS1 protein of influenza virus) at the N-terminus, and a C-
 CC terminal histidine tail. The gene encoding the present protein was cloned
 CC and used to evaluate the effectiveness of tumour-specific antigens.
 CC (Updated on 12-SEP-2003 to standardise OS field)
 XX
 SQ Sequence 262 AA;
 Query Match 54.5%; Score 908; DB 3; Length 262;
 Best Local Similarity 97.2%; Pred. No. 1.7e-90;
 Matches 173; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 69 AVHFSAEWFGVTNNTSYKAFSAARVTARVGLVGLVGINITLTGTPVHQNETIDYNE 128
 Db 76 ALKMTMEFVGTVNTNTSYKAFSAARVTARVGLVGLVGINITLTGTPVHQNETIDYNE 135
 QY 129 QFTWRLKENYAAEYANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYASATLWVAF 188
 Db 136 QFTWRLKENYAAEYANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYASATLWVAF 195
 QY 189 CFWLLSNVLTSTPAPLYGGLALLTTGAFALFGVFALASISSVPLCPRLGSSALLTQY 246
 Db 196 CFWLLSNVLTSTPAPLYGGLALLTTGAFALFGVFALASISSVPLCPRLGSSALLTQY 253
 RESULT 7
 AAB97510
 ID AAB97510 standard; protein; 343 AA.
 XX
 AC AAB97510;
 XX
 DT 27-JUN-2002 (first entry)
 XX
 DE Novel human protein SEQ ID NO: 778.
 XX
 KW Human; antianaemic; vulnery; antiinflammatory; immunomodulator;
 KW antifertility; cerebroprotective; cyostatic; rheumatic; gene therapy;
 KW neuroprotective; antiparkinsonian; protein therapy; EST;
 KM expressed sequence tag.
 XX
 OS Homo sapiens.
 XX
 PN WO200222660-A2.
 XX
 PD 21-MAR-2002.

XX 10-SEP-2001; 2001WO-US026015.
 PF
 XX 11-SEP-2000; 2000US-00659671.
 PR
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
 XX
 DR WPI; 2002-292408/33.
 DR N-PSDB; ABR32696.
 XX
 PT An isolated polynucleotide for treating diseases associated with its
 PT encoded polypeptide such as cancer and multiple sclerosis.
 XX
 PS Claim 20; SEQ ID NO 778; 509pp; English.
 XX
 CC The present invention provides the protein and coding sequences of 444
 CC novel human proteins. These were isolated from expressed sequences tags
 CC (ESTs). They can be used to stimulate cell growth, to regulate
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
 CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
 CC rheumatoid arthritis, and to treat nervous system disorders e.g.
 CC Parkinson's disease. The present sequence is a protein of the invention
 XX
 SQ Sequence 343 AA;
 Query Match 53.0%; Score 883; DB 5; Length 343;
 Best Local Similarity 58.8%; Pred. No. 1.3e-87;
 Matches 173; Conservative 40; Mismatches 73; Indels 8; Gaps 3;
 QY 9 PFYQPRHAAGSVPLLVLFVLAALASFLILPGIRGHSRMFWLVRLSLFICAEIV 68
 Db 9 PFYAGPKPTFPMDTTLASIMIFLTALATFIVILPGIRGKTRLFWLRVVTSLFICAIL 68
 QY 69 AVHFSAEWFGVTNNTSYKAFSAARVTARVGLVGLVGINITLTGTPVHQNETIDYNE 128
 Db 69 AVHFSAEWFGVTNNTSYKAFSAARVTARVGLVGLVGINITLTGTPVHQNETIDYNE 128
 QY 129 QFTWRLKENYAAEYANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYASATLWVAF 188
 Db 129 EFTWRLGENYAEYAKALEKGLPDPVLYLEKETPRSCGLYRQYHLAGHYASATLWVAF 188
 QY 189 CFWLLSNVLTSTPAPLYGGLALLTTGAFALFGV--FALASISSVPLCPRLGSSALLTQY 246
 Db 189 LCVLLANVMLSMPALVYGGVLLATGIFQLALLFSAVATSLTSP--CPHLGASVATHH 247
 QY 247 GAFFWTLATGVLCFLGAVVSLQYVPSALRTILDOSAK-----DCSQERG 295
 Db 248 GPAFWITLTGTLGVLLGLAMAVAHRMQPHRLKAFNQSVDDEPMLWSPSEEG 301
 RESULT 8
 AAB6071
 ID AAB6071 standard; protein; 343 AA.
 XX
 AC AAB6071;
 XX
 DT 30-MAR-2001 (first entry)
 XX
 DE Human INTERCEPT 296.
 XX
 KW TANGO protein; INTERCEPT protein; neurological disorder;
 KW central nervous system; focal brain disorder; bipolar affective disorder;
 KW global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
 KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
 KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
 KW neuropsychiatric; psychoactive substance use; anxiety.

OS	Homo sapiens.
XX	
PN	WO200077239-A2.
XX	
PD	21-DEC-2000.
XX	
PF	24-MAY-2000; 2000WO-US014858.
XX	
PR	14-JUN-1999; 99US-00333159.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Mccarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX	
DR	WPI; 2001-032313/04.
DR	N-PSDB; AAF45133, AAF45134.
XX	
PT	TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT	screening assays and diagnostic assays and for the treatment of
PT	neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT	disease.
XX	
PS	Claim 8; Fig 7; 359pp; English.
XX	
CC	The present invention relates to TANGO or INTERCEPT proteins and coding
CC	sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC	AAB66054-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC	sequences are useful for the treatment of neurological disorders such as
CC	central nervous system (CNS) disorders, CNS-related disorders, focal
CC	brain disorders, global-diffuse cerebral disorders and other neurological
CC	and cerebrovascular disorders. The CNS disorders include Alzheimer's
CC	disease, senile dementia, Huntington's disease, amyotrophic lateral
CC	sclerosis, Parkinson's, Gilles de la Tourette's syndrome, autonomic
CC	function disorders such as hypertension and sleep disorders,
CC	neuropsychiatric disorders, psychoactive substance use disorders,
CC	anxiety, and bipolar affective disorder
XX	
SQ	Sequence 343 AA:
Query Match	52.5%; Score 874; DB 4; Length 343;
Best Local Similarity	58.5%; Pred. No. 1.3e-86;
Matches 172; Conservative	40; Mismatches 74; Indels 8; Gaps 3;
QY	9 PFYPQRRHAAGFSVPLLIVLVFLALAASTLLIPGIRGSRWFVLVRLSLFTGAETV 68
DB	9 PFYAGPKRTPPMDTTLASIIIMIFLTAATFIVILPGIRGRRLFWLDRVTSLFIGAAIL 68
QY	69 AVHFSAEWVFVTNTNITSYKAFSARVTARVGILLVGLGINITLTGPVHQLNETHDYNE 128
DB	69 AVNFSSSESVGVSTINTSYKAFSSSEWISADIGLVGLGVNITLTGTVPVQOLNETINYN 128
QY	129 QFTWRLEKNYAEEYANALEKGLPPDVLTYLAKEFTPSSPCGIHYOHLACHYASATLMVAF 188
DB	129 EFTWRLEGNYYAEECAKALEKGLPPDVLTYLAKEFTTSPCGIXRYQLAGHTSAMLMVAF 188
QY	189 CFWLINSVLLSTPAPLYGGALLTTGAFALFGV--FALASTSSVPLCPURLGSSSALTTOY 246
DB	189 LCWLLANWMLMPVLVYGGYMLLATGIFQLALLLFFSMATSLTSP-CPLHLGASVATHFH 247
QY	247 GAAPFWTLATGVLCEFLGGAVVSIQYVRPSALRTLLDOSAK-----DCSQERGG 295
DB	248 GPAFWITLTTGLLCVLLGLAMAVAHRMQPHRLKAFFNQSDVEDDPMLEWSPEEGG 301
RESULT 9	
ID	ABU08370 standard; protein; 343 AA.
XX	
AC	ABU08370;
XX	
DT	03-JUN-2003 (first entry)
XX	
DE	Amino acid sequence for human INTERCEPT 296.

XX Human; INTERCEPT 296; TANGO; cellular process; tissue typing;
 KW forensic biology; cellular protease activity; cell interaction;
 KW development; blood disorder; haematopoietic cell-related disorder;
 KW growth; cell proliferation; cell differentiation; gamma delta T-cell;
 KW immune system; metabolic disorder; homeostatic disorder; anaemia;
 KW developmental bone disorder; osteoporosis; bacterial infection;
 KW thrombocytopaenia; renal failure; liver disease; cystic fibrosis;
 KW obesity; cancer; bronchitis; asthma; emphysema; pulmonary oedema;
 KW respiratory distress syndrome; osteopathic; antibacterial; antihaemic;
 KW thrombolytic; nephrotropic; antiobesity; hepatotropic; cytostatic;
 KW antiinflammatory; antiasthmatic.
 XX
 OS Homo sapiens.
 PN US2002182675-A1.
 XX
 PD 05-DEC-2002.
 XX
 PF 25-OCT-2001; 2001US-00042431.
 XX
 PR 14-JUN-1999; 99US-00333159.
 XX 24-MAY-2000; 2000US-00578063.
 PA (MILL-) MILLENNIUM PHARM INC.
 PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;
 DR WPI; 2003-328617/31.
 XX N-PSDB; ABX94107.
 PT New TANGO and INTERCEPT proteins, useful as modulating agents in
 PT regulating a variety of cellular processes, in chromosome mapping, in
 PT tissue typing, and in forensic biology.
 XX
 PS Disclosure; Fig 7A-7C; 232pp; English.
 XX
 CC The present invention relates to the isolation of novel TANGO or
 CC INTERCEPT proteins, and the polynucleotide sequences encoding them. The
 CC invention discloses sequences for human TANGO 202, TANGO 234, TANGO 265,
 CC TANGO 273, TANGO 286, TANGO 294 and INTERCEPT 296. The sequences for
 CC murine TANGO 202 and TANGO 273 are also provided. The TANGO polypeptides
 CC and the polynucleotide sequences encoding them are useful as modulating
 CC agents in regulating a variety of cellular processes. The polynucleotide
 CC sequences are useful as primers or hybridisation probes for the detection
 CC of nucleic acids encoding TANGO polypeptides, chromosome mapping, tissue
 CC typing, and in forensic biology. TANGO 202 can be used to diagnose,
 CC prevent or treat disorders relating to aberrant cellular protease
 CC activity, inappropriate interaction of cells with mediators,
 CC inappropriate development, and blood and haematopoietic cell-related
 CC disorders. TANGO 234 can be used to modulate growth, proliferation,
 CC survival, differentiation, and activity of gamma delta T-cells. TANGO 265
 CC can be used to prevent, diagnose and treat disorders characterised by
 CC aberrant organisation or development of a tissue or organ, and for
 CC modulating differentiation of cells of the immune system. TANGO 273 is
 CC useful for diagnosing, treating or preventing e.g. metabolic, homeostatic
 CC and developmental bone disorders (e.g. osteoporosis), and bacterial
 CC infection. TANGO 286 is useful for treating or preventing e.g. anaemia,
 CC thrombocytopaenia, renal failure or liver disease. TANGO 294 is useful
 CC for treating or preventing e.g. cystic fibrosis or obesity. INTERCEPT 296
 CC is useful for diagnosing, treating or preventing e.g. cancers,
 CC bronchitis, cystic fibrosis, asthma, emphysema, pulmonary oedema, or
 CC adult and infant respiratory distress syndromes. The present sequence
 CC represents human INTERCEPT 296
 XX
 SQ Sequence 343 AA;
 Query Match 52.5%; Score 874; DB 6; Length 343;
 Best Local Similarity 58.5%; Pred. No. 1.3e-86;
 Matches 172; Conservative 40; Mismatches 74; Indels 8; Gaps 33

Db 9 PFYAGPKPTFPMDTLLASIMIFLALATFIVILPGIRGKTRLLFWLRRVTSLFIGAAIL 68
QY 69 AVHFSAEWFGVTNNTNTSYKAFSAARVTARVGLVGLGINITLTGTPVHQLNETIDYNE 128
Db 69 AVNFSESWSVGVSTNTSYKAFSSSEWISADIGLVGLGVNITLTGTPVQQLNETININE 128
QY 129 QFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPPSSPCGLYHQYHLAGHYASATLWVAF 188
Db 129 EFTWRLGENYAEECAKALEKGLPDPVLYLAKEFTPPSPCGLYRQYRLAGHYTSAMLWVAF 188
QY 189 CFWLLSNVLLSTPAPLYGGLALLTTGAFALFGV--FALASISSVPLCPRLGSSALLTQY 246
Db 189 LCWLLANWMLSMPLVYGYGMYLLATGIFQLLALLFESMATSLTSP-CPLHLGASVLHTHH 247
QY 247 GAAPWVTLATGVLCLFLGGAIVSLQYVRPSALRTLIDQSAK-----DCSQERGG 295
Db 248 GPAFWITLTGTLGVLLGLAMAVAHRMQPHRLKAFENQSVDEDPMLEWSPREEGG 301

RESULT 10

ABO32725 standard; protein; 343 AA.

ABO32725;

17-SEP-2003 (first entry)

Secreted polypeptide-related protein #100.

Human; TANGO; INTERCEPT: secreted polypeptide; immune disorder;
hormonal disorder; proliferative disorder; cancer; thyroid disorder;
diabetes; multiple sclerosis; lupus; neurological disorder; anaemia;
Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
myocardial infarction; congestive heart disease; blood platelet disorder;
thrombocytopenia; blood vessel; atherosclerosis; vasculitis.

Homo sapiens.

US2003022279-A1.

30-JAN-2003.

12-JAN-2001; 2001US-00759130.

14-JUN-1999; 99US-00333159.
29-JUN-1999; 99US-00342364.
10-SEP-1999; 99US-00393996.
19-OCT-1999; 99US-00420707.
07-JAN-2000; 2000US-00479249.
27-APR-2000; 2000US-00559497.
24-MAY-2000; 2000US-00578063.
16-JUN-2000; 2000US-00596194.
23-JUN-2000; 2000US-00602871.
30-JUN-2000; 2000US-00608452.

(FRAS/) FRASER C C.
(BARN/) BARNES T M.
(SHAR/) SHARP J D.
(KIRS/) KIRST S J.
(MYER/) MYERS P S.
(LEIB/) LEIBY K R.
(HOLT/) HOLTZMAN D A.
(MCCA/) MCCARTHY S A.
(WRIGHT/) WRIGHTON N.
(MACK/) MACKAY C R.
(GOOD/) GOODEARL A D J.

Fraser CC, Barnes TM, Sharp JD, Kirst SJ, Myers PS, Leiby KR;
Holtzman DA, McCarthy SA, Wrighton N, Mackay CR, Goodearl ADJ;
WPI; 2003-456290/43.
N-PSDB; ACD66786, ACD66787.

PT New nucleic acid molecule encoding a secreted protein (e.g. TANGO 202,
PT TANGO 210 or INTERCEPT 217), useful for diagnosing, preventing or
PT treating disorders such as cancer, diabetes or atherosclerosis, and in
PT forensic biology.

PS Claim 9; Fig 30A-30C; 482pp; English.

CC The invention relates to secreted polypeptide-related proteins and
CC nucleic acids (TANGO and INTERCEPT proteins and nucleic acids). The
CC nucleic acids, proteins and antibodies specific to the proteins are
CC useful in screening assays, predictive medicine (e.g. diagnostic assays,
CC prognostic assays, monitoring clinical trials and pharmacogenetics) and
CC prophylactic and therapeutic methods. The sequences are used in
CC diagnosing, preventing or treating proliferative disorders (e.g.
CC cancers), hormonal disorders (e.g. diabetes or thyroid disorders), immune
CC disorders (e.g. multiple sclerosis or lupus), neurological disorders
CC (e.g. Alzheimer's disease or Parkinson's disease), cardiovascular
CC disorders (e.g. myocardial infarction or congestive heart disease), blood
CC platelet disorders (e.g. thrombocytopenia or anaemia) and disorders
CC involving blood vessels (e.g. atherosclerosis or vasculitis). The nucleic
CC acids may also be used in chromosome mapping, tissue typing and forensic
CC biology, and as surrogate markers. This sequence represents a secreted
CC polypeptide-related protein of the invention. Note: The sequence data for
CC this patent was obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html

Sequence 343 AA;

Query Match 52.5%; Score 874; DB 6; Length 343;
Best Local Similarity .58.5%; Pred. No. 1.3e-86;
Matches 172; Conservative 40; Mismatches 74; Indels 8; Gaps 3;

QY 9 PFYQPRHAGFVSPIILVILVFLAASFLILPGIRGSRFWMLVRVLSIFIGAEIV 68
Db 9 PFYAGPKPTFPMDTLLASIMIFLALATFIVILPGIRGKTRLLFWLRRVTSLFIGAAIL 68
QY 69 AVHFSAEWFGVTNNTNTSYKAFSAARVTARVGLVGLGINITLTGTPVHQLNETIDYNE 128
Db 69 AVNFSESWSVGVSTNTSYKAFSSSEWISADIGLVGLGVNITLTGTPVQQLNETININE 128
QY 129 QFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPPSSPCGLYHQYHLAGHYASATLWVAF 188
Db 129 EFTWRLGENYAEECAKALEKGLPDPVLYLAKEFTPPSPCGLYRQYRLAGHYTSAMLWVAF 188
QY 189 CFWLLSNVLLSTPAPLYGGLALLTTGAFALFGV--FALASISSVPLCPRLGSSALLTQY 246
Db 189 LCWLLANWMLSMPLVYGYGMYLLATGIFQLLALLFESMATSLTSP-CPLHLGASVLHTHH 247
QY 247 GAAPWVTLATGVLCLFLGGAIVSLQYVRPSALRTLIDQSAK-----DCSQERGG 295
Db 248 GPAFWITLTGTLGVLLGLAMAVAHRMQPHRLKAFENQSVDEDPMLEWSPREEGG 301

RESULT 11

ADA55266 standard; protein; 298 AA.

ADA55266;

20-NOV-2003 (first entry)

Human protein, SEQ ID 2834.

Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
Gene Therapy; human; secretory protein; membrane proteins; cancer;
inflammatory disease; osteoporosis; neurological disease.

Homo sapiens.

EP1293569-A2.

19-MAR-2003.

PF 21-MAR-2002; 2002EP-00006586.
 XX
 PR 14-SEP-2001; 2001JP-00328381.
 PR 24-JAN-2002; 2002US-0350435P.
 XX
 PA (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
 XX
 DR WPI; 2003-395539/38.
 DR N-PSDB; ADA53627.
 XX
 PT New polynucleotides encoding full-length polypeptides, e.g. secretory
 PT and/or membrane proteins, useful for developing medicines for diseases in
 PT which the gene is involved, or as target molecules for gene therapy.
 XX
 PS Claim 14; SEQ ID NO 2834; 205bp; English.
 XX
 CC The present invention relates to novel human secretory or membrane
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
 CC ADA54071). The coding sequences are useful in the gene therapy of
 CC diseases caused by abnormalities of the proteins, e.g. cancer,
 CC inflammatory diseases, osteoporosis or neurological disease.
 XX
 SQ Sequence 298 AA;

Query Match	41.9%	Score 697.5;	DB 6;	Length 298;
Best Local Similarity	48.6%;	Pred. No. 2.3e-67;		
Matches 143; Conservative	32;	Mismatches 66;	Indels 53;	Gaps 4;

[illegible]

RESULT 12
ABG03520
ID ABG03520 standard; protein; 445 AA.

AC ABG03520;

DT 13-FEB-2002 (first entry)

Novel human diagnostic protein #3511.

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder. KW

OS Homo sapiens.

PN W0200175067-A2.

PD 11-OCT-2001.

XX 3

PF 30-MAR-2001; 2001WO-US008631.

XX

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT,

DR WPI; 2001-639362/73.

DR N-PSDB; AAS67707.

PT New isolated polynucleotide and encoded polypeptides, useful in PT diagnostics, forensics, gene mapping, identification of mutations PT responsible for genetic disorders or other traits and to assess PT biodiversity.

Claim 20; SEQ ID NO 33879; 103pp; English

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences

AA: Sequence 445 AA;

Query Match	27.6%	Score 460.5;	DB 4;	Length 445;
Best Local Similarity	51.1%;	Pred. No. 3.6e-41;		
Matches 91; Conservative	14;	Mismatches 28;	Indels 45;	Gaps 1;

QY 9 PFYPQPRHAGFSVPLLIVLVELALAASELLIPGIRGHSRWFVLVRVLLSLFIGAEIV 68

D5 77 PFYAGPKPTFPMDTLASIMIFLTALATFIVILPGIRGKTRLFWLRRVTSIFGAAIL 130

69 AVHFSAEVFGTVNTNTSYKAFSARVTARVGLLVGLEGINITLTGTPVHQNETIDYNE 128

D6 137 -----GTPVQQLNETINYNNE 155

DQ 129 QFTWRLKENYAEEYANALEKGLPDPVLYLAKEFTTSSPCGJLHYOHLA GHYASATLMW 186
:
DB 152 EFTWRLGENYAEEYAKALEKGLPDPVLYLAKEFTTPRSPCGJLYOYRLAGHTSAMLMW 209

RESULT 13

ID AAB66073 standard; protein; 112 AA.

AC AAB66073;

DT 30-MAR-2001 (first entry)

Human INTERCEPT 296 extracellular domain #2.

KW TANGO protein; INTERCEPT protein; neurological disorder;

global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease; KW

KW senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KW Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KW neuropsychiatric; psychoactive substance use; anxiety.
OS Homo sapiens.
XX
PN WO200077239-A2.
XX
PD 21-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US014858.
XX
PR 14-JUN-1999; 99US-00333159.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI McCarthy SA, Fraser CC, Sharp JD, Barnes TM;
XX
DR WPI; 2001-032313/04.
XX
XX TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT screening assays and diagnostic assays and for the treatment of
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT disease.
XX
PS Claim 8; Page 331; 359pp; English.
XX
CC The present invention relates to TANGO or INTERCEPT proteins and coding
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC sequences are useful for the treatment of neurological disorders such as
CC central nervous system (CNS) disorders, CNS-related disorders, focal
CC brain disorders, global-diffuse cerebral disorders and other neurological
CC and cerebrovascular disorders. The CNS disorders include Alzheimer's
CC disease, senile dementia, Huntington's disease, amyotrophic lateral
CC sclerosis, Parkinson's, Gilles de la Tourette's syndrome, autonomic
CC function disorders such as hypertension and sleep disorders,
CC neuropsychiatric disorders, psychoactive substance use disorders,
CC anxiety, and bipolar affective disorder
XX
SQ Sequence 112 AA;

Query Match 27.5%; Score 458; DB 4; Length 112;
Best Local Similarity 76.8%; Pred. No. 9.3e-42;
Matches 86; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

QY 71 HFSAEWFGVTNTNTSYKAFSARVTARVGLVLEGINITLTGTPVHQNETIDYNEQF 130
Db 1 NFSSEWSVGQVSTNTSYKAFSSEWISADIGLVGQVNTLTGTPVQQLNETINNEEF 60

QY 131 TWRLKENYAAEYANALEKGLPDPVLYLAEKFTPPSSPCGLYHQYHLAGHYASA 182
Db 61 TWRLGENYAECAKALEKGLPDPVLYLAEKFTPPSSPCGLYRQYRLAGHYTSA 112

RESULT 14
ABO32727
ID ABO32727 standard; protein; 112 AA.
XX
AC ABO32727;
XX
DT 17-SEP-2003 (first entry)
XX
DE Secreted polypeptide-related protein #101.
XX
KW Human; TANGO; INTERCEPT; secreted polypeptide; immune disorder;
KW hormonal disorder; proliferative disorder; cancer; thyroid disorder;
KW diabetes; multiple sclerosis; lupus; neurological disorder; anaemia;
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KW myocardial infarction; congestive heart disease; blood platelet disorder;
KW thrombocytopenia; blood vessel; atherosclerosis; vasculitis.
XX
OS Homo sapiens.

XX US2003022279-A1.
PN
XX
PD 30-JAN-2003.
XX
PF 12-JAN-2001; 2001US-00759130.
XX
PR 14-JUN-1999; 99US-00333159.
PR 29-JUN-1999; 99US-00342364.
PR 10-SEP-1999; 99US-00393996.
PR 19-OCT-1999; 99US-00420707.
PR 07-JAN-2000; 2000US-00479249.
PR 27-APR-2000; 2000US-00559497.
PR 24-MAY-2000; 2000US-00578063.
PR 16-JUN-2000; 2000US-00596194.
PR 23-JUN-2000; 2000US-00602871.
PR 30-JUN-2000; 2000US-00608452.
XX
PA (FRAS/) FRASER C C.
PA (BARN/) BARNES T M.
PA (SHAR/) SHARP J D.
PA (KIRS/) KIRST S J.
PA (MYER/) MYERS P S.
PA (LEIB/) LEIBY K R.
PA (HOLT/) HOLTZMAN D A.
PA (MCCA/) MCCARTHY S A.
PA (WRIG/) WRIGHTON N.
PA (MACK/) MACKAY C R.
PA (GOOD/) GOODEARL A D J.
XX
PI Fraser CC, Barnes TM, Sharp JD, Kirst SJ, Myers PS, Leiby KR;
PI Holtzman DA, McCarthy SA, Wrighton N, Mackay CR, Goodearl ADJ;
XX
DR WPI; 2003-456290/43.
XX
XX New nucleic acid molecule encoding a secreted protein (e.g. TANGO 202,
PT TANGO 210 or INTERCEPT 217), useful for diagnosing, preventing or
PT treating disorders such as cancer, diabetes or atherosclerosis, and in
PT forensic biology.
XX
PS Claim 9; SEQ ID NO 427; 482pp; English.
XX
CC The invention relates to secreted polypeptide-related proteins and
CC nucleic acids (TANGO and INTERCEPT proteins and nucleic acids). The
CC nucleic acids, proteins and antibodies specific to the proteins are
CC useful in screening assays, predictive medicine (e.g. diagnostic assays,
CC prognostic assays, monitoring clinical trials and pharmacogenetics) and
CC prophylactic and therapeutic methods. The sequences are used in
CC diagnosing, preventing or treating proliferative disorders (e.g.
CC cancers), hormonal disorders (e.g. diabetes or thyroid disorders), immune
CC disorders (e.g. multiple sclerosis or lupus), neurological disorders
CC (e.g. Alzheimer's disease or Parkinson's disease), cardiovascular
CC disorders (e.g. myocardial infarction or congestive heart disease), blood
CC platelet disorders (e.g. thrombocytopenia or anaemia) and disorders
CC involving blood vessels (e.g. atherosclerosis or vasculitis). The nucleic
CC acids may also be used in chromosome mapping, tissue typing and forensic
CC biology, and as surrogate markers. This sequence represents a secreted
CC polypeptide-related protein of the invention. Note: The sequence data for
CC this patent was obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 112 AA;

Query Match 27.5%; Score 458; DB 6; Length 112;
Best Local Similarity 76.8%; Pred. No. 9.3e-42;
Matches 86; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

QY 71 HFSAEWFGVTNTNTSYKAFSARVTARVGLVLEGINITLTGTPVHQNETIDYNEQF 130
Db 1 NFSSEWSVGQVSTNTSYKAFSSEWISADIGLVGQVNTLTGTPVQQLNETINNEEF 60

QY 131 TWRLKENYAAEYANALEKGLPDPVLYLAEKFTPPSSPCGLYHQYHLAGHYASA 182
Db 61 TWRLKENYAAEYANALEKGLPDPVLYLAEKFTPPSSPCGLYHQYHLAGHYASA 182

Db 61 TWRLGENYAECAKALEKGLPDPVLYIAEKFTPRSPCGLYRQYRLAGHYTSA 112

RESULT 15

AAB66060
ID AAB66060 standard; protein; 383 AA.

XX AAB66060;

DT 30-MAR-2001 (first entry)

DE C06E1.3 related protein, CRP.

KM Gene therapy; TANGO protein; INTERCEPT protein; neurological disorder;
KM central nervous system; focal brain disorder; bipolar affective disorder;
KM global-diffuse cerebral disorder; cerebrovascular; Alzheimer's disease;
KM senile dementia; Huntington's disease; amyotrophic lateral sclerosis;
KM Parkinson's; Gilles de la Tourette's syndrome; hypertension; sleep;
KM neuropsychiatric; psychoactive substance use; anxiety.

XX Caenorhabditis elegans.

PN WO200077239-A2.

PD 21-DEC-2000.

PF 24-MAY-2000; 2000WO-US014858.

PR 14-JUN-1999; 99US-00333159.

PA (MILL-) MILLENNIUM PHARM INC.

PI Mccarthy SA, Fraser CC, Sharp JD, Barnes TM;

DR WPI; 2001-032313/04.

PT TANGO and INTERCEPT nucleic acids, proteins, and antibodies, useful for
PT screening assays and diagnostic assays and for the treatment of
PT neurological diseases such as Alzheimer's, Parkinson's and Huntington's
PT disease.

PS Claim 8; Fig 7; 359pp; English.

XX The present invention relates to TANGO or INTERCEPT proteins and coding
CC sequences (see AAF45121-F45136 and AAF45138-F45139 and AAB66031-B66057,
CC AAB66064-B66083 and AAB66085). The TANGO/INTERCEPT proteins and coding
CC sequences are useful for the treatment of neurological disorders such as
CC central nervous system (CNS) disorders, CNS-related disorders, focal
CC brain disorders, global-diffuse cerebral disorders and other neurological
CC and cerebrovascular disorders. The CNS disorders include Alzheimer's
CC disease, senile dementia, Huntington's disease, amyotrophic lateral
CC sclerosis, Parkinson's, Gilles de la Tourette's syndrome, autonomic
CC function disorders such as hypertension and sleep disorders,
CC neuropsychiatric disorders, psychoactive substance use disorders,
CC anxiety, and bipolar affective disorder. The present sequence is a
CC sequence used in a sequence homology alignment with the TANGO/INTERCEPT
CC sequences of the present invention

XX Sequence 383 AA;

Query Match 16.6%; Score 276.5; DB 4; Length 383;
Best Local Similarity 29.9%; Pred. No. 4.2e-21;
Matches 83; Conservative 43; Mismatches 125; Indels 27; Gaps 8;

QY 27 VILVFLALASFLILPGIRGSRMFWLVVLSLFTGAEIVAVHFSAEWFGVTVNTS 86

Db 14 IFSVFLIPLIAYLILPGVR-RKRVVTYVYVLMVAVGALLIASLIYPCWASGSQMITYQ 72

QY 87 YKAFSARVTAHVGLLVLEGINITLTGTPTVHQLNETID-----YNEQFTWRLKENY 138

Db 73 FRGHSNERILAKIGVEIGLOKYNVTLKFERLLSSNDVLPQSDMTLEYNEGFDISGTS 132

QY 139 AAEYANALEKGLDPPVLYIAEKFT---PSSPCGLYHQYHLAGHYASATLVWAFCEWLLSN 195

Db 133 AEALHHGLENGLPYPMLSVLEYPFLNQDSFDWG--RHYRVAGHYTHAAIWFAFACWCLSV 190
QY 196 VL-LSTPAFLYGGIALTTGAFALFGVFALASISSVPLCPRLGSSA-----LTTQYG 247
Db 191 VLMLFLPHNAV--KSILATGISCLIACLVYLLISP---CELRIFTGENFERVDLTATFS 245
QY 248 AAFWVTLATGVLCFLPGAVVSLQYVRPSALRTLIDQS 285
Db 246 FCFYLIFAIIGILCVLGLGIGICEHWRIYTLSTFLDAS 283

Search completed: February 23, 2004, 19:16:58
Job time : 103 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:18:06 ; Search time 44 Seconds
(without alignments)
375.462 Million cell updates/sec

Title: US-09-936-456-2
Perfect score: 1666
Sequence: 1 MTLWMGVLPFYQPRHAAGF.....GDPLHKQALPDLKCTITNL 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	109.5	6.6	457	US-09-543-681A-6481	Sequence 6481, Ap
2	108	6.5	660	US-09-489-039A-8557	Sequence 8557, Ap
3	104	6.2	398	US-09-489-039A-8645	Sequence 8645, Ap
4	99.5	6.0	533	US-09-328-352-7892	Sequence 7892, Ap
5	98.5	5.9	317	US-09-305-856B-12	Sequence 12, Appl
6	98.5	5.9	416	US-09-489-039A-9296	Sequence 9296, Ap
7	97.5	5.9	432	US-08-677-049-8	Sequence 8, Appl
8	97	5.8	413	US-09-489-039A-7562	Sequence 7562, Ap
9	96.5	5.8	429	US-09-252-991A-19571	Sequence 19571, A
10	96	5.8	446	US-09-489-039A-13214	Sequence 13214, A
11	96	5.8	446	US-09-543-681A-6579	Sequence 6579, Ap
12	95	5.7	359	US-08-307-382-2	Sequence 2, Appl
13	95	5.7	359	US-08-366-779-2	Sequence 2, Appl
14	95	5.7	359	US-08-478-727-2	Sequence 2, Appl
15	95	5.7	359	US-08-473-508-2	Sequence 2, Appl
16	95	5.7	359	US-08-789-936-2	Sequence 2, Appl
17	95	5.7	359	US-08-833-610-6	Sequence 6, Appl
18	95	5.7	359	US-08-834-033A-16	Sequence 16, Appl
19	95	5.7	359	US-08-934-254-2	Sequence 2, Appl
20	95	5.7	359	US-09-377-452-6	Sequence 6, Appl
21	95	5.7	359	US-09-685-775-2	Sequence 2, Appl
22	94.5	5.7	403	US-09-328-352-7791	Sequence 7791, Ap
23	94.5	5.7	455	US-09-134-001C-5597	Sequence 5597, Ap
24	94	5.6	1160	US-09-328-352-6826	Sequence 6826, Ap
25	93.5	5.6	569	US-09-252-991A-22870	Sequence 22870, A
26	92.5	5.6	421	US-09-489-039A-7699	Sequence 7699, Ap
27	92.5	5.6	467	US-09-332-041-3	Sequence 3, Appl

28	91.5	5.5	515	4	US-09-328-352-7312	Sequence 7312, Ap
29	91	5.5	437	4	US-09-134-001C-4808	Sequence 4808, Ap
30	90	5.4	424	4	US-09-543-681A-4762	Sequence 4762, Ap
31	90	5.4	503	4	US-09-252-991A-32663	Sequence 32663, A
32	89.5	5.4	415	4	US-09-543-681A-6292	Sequence 6292, Ap
33	89	5.3	441	4	US-09-252-991A-17727	Sequence 17727, A
34	89	5.3	484	4	US-09-489-039A-7690	Sequence 7690, Ap
35	89	5.3	510	4	US-09-252-991A-26663	Sequence 26663, A
36	89	5.3	597	4	US-09-252-991A-24252	Sequence 24252, A
37	89	5.3	909	4	US-09-252-991A-30503	Sequence 30503, A
38	88.5	5.3	455	4	US-09-252-991A-23745	Sequence 23745, A
39	87	5.2	475	4	US-09-328-352-5706	Sequence 5706, Ap
40	87	5.2	482	4	US-09-489-039A-9909	Sequence 9909, Ap
41	87	5.2	485	4	US-09-489-039A-9448	Sequence 9448, Ap
42	86.5	5.2	288	4	US-09-543-681A-5375	Sequence 5375, Ap
43	86.5	5.2	1876	2	US-08-619-554-2	Sequence 2, Appl
44	86	5.2	450	4	US-09-252-991A-25019	Sequence 25019, A
45	86	5.2	578	4	US-09-489-039A-14097	Sequence 14097, A

ALIGNMENTS

RESULT 1
US-09-543-681A-6481
; Sequence 6481, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIR
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543, 681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6481
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6481

Query Match 6.6%; Score 109.5; DB 4; Length 457;
Best Local Similarity 20.2%; Pred. No. 0.0055;
Matches 68; Conservative 39; Mismatches 113; Indels 117; Gaps 15;

QY	20	FSVPLIVILVFLALASFLLLPGIRHSRW-----FWLVRLSLFAG	66
DB	23	FWSEHIIIPAITIGLAFLLLEGL-----WLKTEPSYLSLYQWIKIFAVNFGMGV	77
QY	67	---IVAVHFAEW---FVGTV-NTNTSYKAFSAKAVTA-----RVGLLVGLE	106
DB	78	SGLVMAYQFGTNWSGFSQAGSITGPMUYEVLTAFPLEAGFLGIMLFGMNKVGKLFHF	137
QY	107	GINITLTGPVHQINETIDYNEQFTWRKEN--YAAEYANALEKGLPDPVLYLAKEFTPS	164
DB	138	ATCMVALGTLISTF-----WISSNSWMQTPQGYAIENGVVVPDMLQYVFNPS	186
QY	165	SPCGLYHQVHLAGHYASATLVAF-C-FWLISNVLLSTPAPLYGGLALLTTGAFALFVFA	223
DB	187	FVYRLIHM--SIGAFLASLFTASCAAML-----	214
QY	224	IASISSVPLCPRLGSSALTTQYGAFAVVTLATGVLCFLGGAIVSLQYVRPSALFTLD	283
DB	215	---LKGQNTAPVR-----KMFSMALMLVLIAPIAFVGDA-----HGNTLLEH	255
QY	284	QSAKDCSQE-----RGSSPLILGDPHLKQALPDL 313	
DB	256	QPAKIAIEGHWDNATKEATPLIL-----FGIPDM 285	

RESULT 2

US-09-489-039A-8557
; Sequence 8557, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8557
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8557

Query Match 6.5%; Score 108; DB 4; Length 660;
Best Local Similarity 23.0%; Pred. No. 0.014;
Matches 64; Conservative 39; Mismatches 105; Indels 70; Gaps 15;

QY 65 AEIVAVH-----FSAEMFVGTNTNTSYKAFSAARVTARVGLVG-----LEGINITLT 113
DB 161 ARIGAIHSVIFGFSPEAVAGRIIDSSRLV-----ITADEGLRAGRAIPLKXVDDALK 215
QY 114 GTPVHQLNE-----TIDYNE--QFTWR-LKENYAAYANALEKGLPDPVLYLAEKF 161
DB 216 NPNVKSIEHYVVLKRTGNDWQGRDLWSDLIANAASEH-RPYEMNAEDPLIYLTSG 274
QY 162 TPSSPCGLYHQYHLAGH--YASATLWYAF-----CFWLLSNVLTSTPAPLYGIALLT 213
DB 275 STGKPKGVLHT--TGGYLVYAATTFKYVFDYHPGDIYWCADV-----GWTG 320
QY 214 GAPALFGVFALASISSVPLCPLRLGSSALTTOYGAFWVTLATGVLCFLGAVVSLQYV 273
DB 321 HSYLLYG-----PLACGATLMEB-GVPMWPTPAR--MCQVVDKHKVSIPLYT 364
QY 274 RPSALRTLDDQSAKDCSQERGGSPIL---GDPLHKQA 308
DB 365 APTAIRALMAGDKAIEGTRSSLRILGSVGEPIPEA 402

RESULT 3

US-09-489-039A-8645
; Sequence 8645, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8645
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8645

Query Match 6.2%; Score 104; DB 4; Length 398;
Best Local Similarity 21.8%; Pred. No. 0.017;
Matches 71; Conservative 39; Mismatches 94; Indels 122; Gaps 15;

QY 12 PQRHAA-----GFSVPLLIVLFLALASFLILLPGIRGSRWFWLV--RV 57
DB 135 PEIRGAAYGLRQALDTVGAFGLGLLAVLLMFI-WANDFHAI-----FWAVIPIAV 183
QY 58 LLSLFIG-----AEIVAVHFSAEWFGVTNTNTSYKAFSAARVTAR 98

DB 184 LSIILLGFLOEPKSAIAHKRSNPLKRENLKLSAAVWVVAIGSIFTLARFSEAFVLVR 243
QY 99 -----VGLLVGEGINITLTGTPVHQLNETIDYNEQFTWRLEKENYAAYANALEK 148
DB 244 AQOMEIPLTIPLVMVAMNLVSLTAYPFGKLSDSMSHSLKLQWLL----- 290
QY 149 GLDPVLYLAEKFTPSBPCGLYHQYHLAGHYASATLWVAFCFW-----LTSNVLIS 199
DB 291 -----VLIADIVLA-----LSGHW--STLLIGVALWGIHWGMTQGLLAAMVAH 332
QY 200 TPAPLYGIALLLTGAFALFGVFALASISSVPLCPLRLGSSALTTOYGA-AFWVTIATGV 258
DB 333 TAPPELRGTA-----FGMFNIMS-----GLALLASTGAGVLTETF--GA 370
QY 259 LCLFLGAVVSL-----QYVRPSALR 279
DB 371 ASTFYAGAIICVTLIGMRWMPSPAYR 396

RESULT 4

US-09-328-352-7892
; Sequence 7892, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACT
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7892
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7892

Query Match 6.0%; Score 99.5; DB 4; Length 533;
Best Local Similarity 21.9%; Pred. No. 0.079;
Matches 61; Conservative 34; Mismatches 95; Indels 89; Gaps 13;

QY 22 VPLLIVLVLALASFLILLPGIRGSRWFWLVRLSLFICA-----EIVAVHS- 73
DB 32 LPLIFPLLIALFSGAVFSFAL---APYWWWLAILSPALLYATLHNSAKQAFAGWSY 87
QY 74 --AEWFGVTNTNTSY-----KAFSAARVTARVGLLVGEGINITLTGTPVHQLNETID 125
DB 88 GFGLMFVGAFWLYTSHVYGDINAFLSVCMIAVMALVMGL----- 127
QY 126 YNEQFTWRLEKENYAAYANALEKGLPDPVLYLAEKFTPSBPCGLYHQYHLAGHYASATLV 185
DB 128 FTAFTW-----IYRFFPERP-----LTFAPLM 151
QY 186 VAFCEWLLSNVLTSTPAPLYGIAL--LTTGAFALFGVFALASISSVPLCPLRLGSSAL 242
DB 152 IIF-EWAKTWVFTGFPW-LFVGYAFTERLDDGYAPLFGIYVSVIVIALCAL-----V 203
QY 243 TTQYGAAFWVTLATGVLCFLGAVVSLQYVRPSALRTL 281
DB 204 EVLRKRIFWV-IPALLVLGAMGASY-IQFVQPKAAKPL 240

RESULT 5

US-09-305-856B-12
; Sequence 12, Application US/09305856B
; Patent No. 6479236
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; APPLICANT: Galvin, Margaret
; TITLE OF INVENTION: Genotyping the Human
; TITLE OF INVENTION: UDP-glucuronosyltransferase 1 (UGT1) Gene
; FILE REFERENCE: 4389-7 (formerly SEQ-17C1P)

; CURRENT APPLICATION NUMBER: US/09/305,856B
 ; CURRENT FILING DATE: 1999-05-05
 ; PRIOR APPLICATION NUMBER: 60/084,807
 ; PRIOR FILING DATE: 1998-05-07
 ; NUMBER OF SEQ ID NOS: 124
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 12
 ; LENGTH: 317
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-305-856B-12

Query Match 5.9%; Score 98.5; DB 4; Length 317;
 Best Local Similarity 22.9%; Pred.No. 0.047;
 Matches 62; Conservative 29; Mismatches 77; Indels 103; Gaps 15;

QY 4 WNGVLPFYPOPRHAGFSVPLLVFLALASFLILPGIRGSRWFWLVLSLFI 63
 Db 6 WTGLLPY-----VCLLLTCALPRSGKLLVPMMDG-SHMTMQSVYEKLIL 50
 QY 64 -GAELVAHFSAEWFG-----TVNT-NTSY-----KAFSARVTARVGLVLEG 107
 Db 51 RGHVVVWVMEVSWQLGRSLNCTVKTYSYTLLEDQREFWVFADARWTAPLSAFL-- 108
 QY 108 INITLTGTPVHQL-----NETIDYNEQFTWRLKEN-----YAAE 141
 Db 109 --LTSSNGIFDLFNSNCRSLFNDRKLEYE-----LKESCDAVFLDPFRCGLIVAK 159
 QY 142 Y-----ANALEKG--LPDPVLYLAEKFTPSPCGLY-----HQVLAGHYA 180
 Db 160 YFSLPSVVFARGIFCHYLEBGAQCPCAPISYVPRLLIGFSDAMTFEXERVMNHIMLEEH-- 217
 QY 181 SATLWVAFCEWLLSNVL-----LSTPAPLY 205
 Db 218 -----LFCPYFEKNVLEIASEILQTPVTAY 242

RESULT 6

; US-09-489-039A-9296
 ; Sequence 9296, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709,2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 9296
 ; LENGTH: 416
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae
 ; US-09-489-039A-9296

Query Match 5.9%; Score 98.5; DB 4; Length 416;
 Best Local Similarity 22.7%; Pred.No. 0.07;
 Matches 75; Conservative 50; Mismatches 140; Indels 65; Gaps 16;

QY 17 AAGFSVPLLVFLALASFLILPGIRGSR-WFWLVVLSLFI-----GAELVAH 71
 Db 44 AVGLIGTLFVGALKAVAPVLMVMASIANHOGKTSIRPIFLVLLGTFSALTAVL 103
 QY 72 FSAEMFVGTNTNTSYKAFSARVTARVGLVLEGINITLTGTPVHQLNETIDYNEQFT 131
 Db 104 FSF-LFPSTLHLLT-----AADSITPPSGIVEVLRLGLMSWVSNPIDALINA-NYIGLIV 156
 QY 132 WRLKENYAAEYANALEKGLPDPVLY-----LAEKTPSSPCGLYHQ-----YHLAGH 178
 Db 157 WAVGLGFALRHGNDTTKNLINDVSHAVTFIVKVVIRFAPLGI FGLVSSTLATGTFETLWG 216

QY 179 YASATLWVAFCEWLLS---NVLL-----STPAPLYGGLALLT-----TGAFALFGVAFALA 225
 Db 217 YAOQLLVVGCMLLVALLVALVINPLLVFWKIRRNYPPL-----VLTCLRESGVYAFTRSSAA 271
 QY 226 SIS-SVPLC-----PLRLGSSALTTQYGAFFWYTLATGVLCIFLGAVVSIQYVR 274
 Db 272 NIPVNMALCEKXNLDRDTYSVSIPLGATINMAGAAITITVLT-----LAAVHTLNI PVDL 326
 QY 275 PSALRTLDDQSAKDC---SQERGSPLIL 300
 Db 327 PTLA--LLSVASLACAGAGVAGSLLLI 354

RESULT 7

; US-08-677-049-8
 ; Sequence 8, Application US/08677049
 ; Patent No. 5858707
 ; GENERAL INFORMATION:
 ; APPLICANT: Guimaraes, M. Jorge
 ; APPLICANT: Bazan, J. Fernando
 ; APPLICANT: McClanahan, Terrill K.
 ; APPLICANT: Zlotnick, Albert
 ; TITLE OF INVENTION: PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;
 ; TITLE OF INVENTION: NUCLEIC ACIDS; ANTIBODIES
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: DNAX Research Institute
 ; STREET: 901 California Avenue
 ; CITY: Palo Alto
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94304-1104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/677,049
 ; FILING DATE: 03-JUL-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/000,788
 ; FILING DATE: 03-JUL-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ching, Edwin P.
 ; REGISTRATION NUMBER: 34,090
 ; REFERENCE/DOCKET NUMBER: DX0511
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-852-9196
 ; TELEFAX: 415-496-1200
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 432 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; FEATURE:
 ; NAME/KEY: Region
 ; LOCATION: 115...144
 ; OTHER INFORMATION: /note= "Encompasses TM 4 of Figure
 ; FEATURE:
 ; NAME/KEY: Region
 ; LOCATION: 325...359
 ; OTHER INFORMATION: /note= "Encompasses TM 9 of Figure
 ; FEATURE:
 ; NAME/KEY: Region
 ; LOCATION: 363...390
 ; OTHER INFORMATION: /note= "Encompasses TM 10 of Figure
 ; OTHER INFORMATION: 4"

RESULT 13

US-08-366-779-2
Sequence 2, Application US/08366779
Patent No. 5614393
GENERAL INFORMATION:
APPLICANT: Thomas, Terry L.
APPLICANT: Reddy, Avutu S.
APPLICANT: Nuccio, Michael
APPLICANT: Freysinet, Georges L.
APPLICANT: Nunberg, Andrew N.
TITLE OF INVENTION: PRODUCTION OF GAMMA LINOLENIC ACID BY A
TITLE OF INVENTION: DELTA 6-DESATURASE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: United States
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/366,779
FILING DATE: 30-DEC-1994
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Presser, Leopold
REGISTRATION NUMBER: 19,827
REFERENCE/DOCKET NUMBER: 8383ZYXW
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 359 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-366-779-2

Query Match 5.7%; Score 95; DB 1; Length 359;

Best Local Similarity 21.3%; Pred. No. 0.13; Mismatches 101; Indels 74; Gaps 16;
Matches 61; Conservative 50; Mismatches 101; Indels 74; Gaps 16;

QY 26 IVILVFLALASFLILPGIRGSRWFLVRLSLFIGAEIVAVHFSAEWFGVTNTNT 85
DB 45 LIIIVLWLSAWAFVLPAPVI-----FPVRL-----GCVLAIALAA--FSFVGHDA 90
QY 86 SYKAFSA-ARVTARVGLVLEGINITLTGTPVHQINETIDYNEQFTWRLKENYAAE-YA 143
DB 91 NHNAVSSNPHINRVIGMTYDFVGLS-----SFLWRYRHNLYLHHTYT 131
QY 144 NALEKGLP---DPVLYIAEKFTPSPCGLY--HOYHLAGHYASATLWVAFCTWLLSNVLL 198
DB 132 NILGHVDEIHGDGAVRMSPE--QEHVGIRFOQFYIWGLY---LFIFP-YWFLYDVYL 183
QY 199 STPAPLYGGLALLTTGAF-----ALFGVFALASISSVPL-----CPLRLGSSALLTQ 245
DB 184 -----VLNKGKYHDHKIPFPQPLELASLIGIKLWLGYVFGPLALGFSIPEVL 232
QY 246 YGAAP-WTLATGVLCF-LGGAVVSLQYVRPSALRTLDDQSAKDC 289
DB 233 IGASVTWYTYGIWCTIFMLAHVLESTEFLLTPDGESGAIDDEMAIC 278

RESULT 14
US-08-478-727-2

Sequence 2, Application US/08478727

Patent No. 5663068
GENERAL INFORMATION:
APPLICANT: Thomas, Terry L.
APPLICANT: Reddy, Avutu S.
APPLICANT: Nuccio, Michael
APPLICANT: Freysinet, Georges L.
TITLE OF INVENTION: PRODUCTION OF GAMMA LINOLENIC
TITLE OF INVENTION: ACID BY A DELTA 6-DESATURASE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: United States
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,727
FILING DATE: 07-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/307,382
FILING DATE: 14-SEP-1994
APPLICATION NUMBER: US 07/817,919
FILING DATE: 08-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: McNulty, William E.
REGISTRATION NUMBER: 22,606
REFERENCE/DOCKET NUMBER: 8383Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 359 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-727-2

Query Match 5.7%; Score 95; DB 1; Length 359;

Best Local Similarity 21.3%; Pred. No. 0.13; Mismatches 101; Indels 74; Gaps 16;
Matches 61; Conservative 50; Mismatches 101; Indels 74; Gaps 16;

QY 26 IVILVFLALASFLILPGIRGSRWFLVRLSLFIGAEIVAVHFSAEWFGVTNTNT 85
DB 45 LIIIVLWLSAWAFVLPAPVI-----FPVRL-----GCVLAIALAA--FSFVGHDA 90
QY 86 SYKAFSA-ARVTARVGLVLEGINITLTGTPVHQINETIDYNEQFTWRLKENYAAE-YA 143
DB 91 NHNAVSSNPHINRVIGMTYDFVGLS-----SFLWRYRHNLYLHHTYT 131
QY 144 NALEKGLP---DPVLYIAEKFTPSPCGLY--HOYHLAGHYASATLWVAFCTWLLSNVLL 198
DB 132 NILGHVDEIHGDGAVRMSPE--QEHVGIRFOQFYIWGLY---LFIFP-YWFLYDVYL 183
QY 199 STPAPLYGGLALLTTGAF-----ALFGVFALASISSVPL-----CPLRLGSSALLTQ 245
DB 184 -----VLNKGKYHDHKIPFPQPLELASLIGIKLWLGYVFGPLALGFSIPEVL 232
QY 246 YGAAP-WTLATGVLCF-LGGAVVSLQYVRPSALRTLDDQSAKDC 289
DB 233 IGASVTWYTYGIWCTIFMLAHVLESTEFLLTPDGESGAIDDEMAIC 278

RESULT 15

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:18:41 ; Search time 626 Seconds
(without alignments)
107.938 Million cell updates/sec

Title: US-09-936-456-2
Perfect score: 1666
Sequence: 1 MTLWNGVLFFYPQPRHAGF.....GDPLHKQALPDLKCTITNL 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 21153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1666	100.0	320	14 US-10-187-657-1	Sequence 1, Appl1
2	1660	99.6	320	15 US-10-264-237-2502	Sequence 2502, Ap
3	874	52.5	343	10 US-09-759-130B-425	Sequence 425, App
4	874	52.5	343	13 US-10-042-431-55	Sequence 55, Appl
5	697.5	41.9	298	15 US-10-094-749-2834	Sequence 2834, Ap
6	458	27.5	112	10 US-09-759-130B-427	Sequence 427, App
7	458	27.5	112	13 US-10-042-431-57	Sequence 57, Appl
8	276.5	16.6	383	10 US-09-759-130B-410	Sequence 410, App
9	276.5	16.6	383	13 US-10-042-431-40	Sequence 40, Appl
10	113.5	6.2	792	14 US-10-156-761-9442	Sequence 9442, Ap
11	104	6.2	652	9 US-09-815-242-13816	Sequence 13816, A
12	103.5	6.2	502	14 US-10-156-761-8917	Sequence 8917, Ap
13	103	6.2	98	14 US-10-106-698-7554	Sequence 7554, Ap
14	103	6.2	460	15 US-10-369-493-15557	Sequence 15557, A
15	103	6.2	460	15 US-10-369-493-15930	Sequence 15930, A

16	103	6.2	460	15 US-10-369-493-16298	Sequence 16298, A
17	101	6.1	735	15 US-10-369-493-8229	Sequence 8229, Ap
18	99.5	6.0	477	9 US-09-815-242-13840	Sequence 13840, A
19	98.5	5.9	317	9 US-09-305-856B-12	Sequence 12, Appl
20	98.5	5.9	317	14 US-10-247-159-12	Sequence 12, Appl
21	98.5	5.9	464	14 US-10-238-075-709	Sequence 709, App
22	98	5.9	463	9 US-09-815-242-10144	Sequence 10144, A
23	96	5.8	430	14 US-10-080-170-33	Sequence 33, Appl
24	95	5.7	359	13 US-10-029-756-2	Sequence 2, Appl1
25	95	5.7	359	14 US-10-340-779A-12	Sequence 12, Appl
26	95	5.7	359	15 US-10-369-493-2766	Sequence 2766, Ap
27	94.5	5.7	331	9 US-09-738-626-3900	Sequence 3900, Ap
28	94.5	5.7	459	15 US-10-369-493-18271	Sequence 18271, A
29	94	5.6	484	14 US-10-156-761-13878	Sequence 13878, A
30	94	5.6	652	9 US-09-815-242-10449	Sequence 10449, A
31	94	5.6	652	15 US-10-369-493-23658	Sequence 23658, A
32	93.5	5.6	548	14 US-10-289-161A-9	Sequence 9, Appl1
33	93.5	5.6	905	15 US-10-369-493-22577	Sequence 22577, A
34	92.5	5.6	467	9 US-09-887-038-3	Sequence 3, Appl1
35	92.5	5.6	467	10 US-09-828-173-3	Sequence 3, Appl1
36	92.5	5.6	467	14 US-10-410-432-3	Sequence 3, Appl1
37	92.5	5.6	467	15 US-10-310-154-653	Sequence 653, App
38	91.5	5.5	477	11 US-09-833-245-80	Sequence 80, Appl
39	91.5	5.5	477	15 US-10-264-237-2729	Sequence 2729, Ap
40	91.5	5.5	951	14 US-10-276-340-3	Sequence 3, Appl1
41	91.5	5.5	161	14 US-10-170-102-4	Sequence 4, Appl1
42	91	5.5	306	14 US-10-156-761-12838	Sequence 12838, A
43	91	5.5	371	14 US-10-156-761-14576	Sequence 14576, A
44	91	5.5	469	14 US-10-156-761-12877	Sequence 12877, A
45	90.5	5.4	445	14 US-10-275-554-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-10-187-657-1
; Sequence 1, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Iasek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030068311A1 4901066CD1
US-10-187-657-1

Query Match 100.0%; Score 1666; DB 14; Length 320;
Best Local Similarity 100.0%; Pred. No. 1.66-155;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MTLWNGVLFFYPQPRHAGFSVPLILVILVFLAALASFLILLPGIRGHSRMFWLVRLLS	60
Db	1	MTLWNGVLFFYPQPRHAGFSVPLILVILVFLAALASFLILLPGIRGHSRMFWLVRLLS	60
QY	61	LFIGAEIVAVHFSAEWFFVGTANTNTSYKAFSARVTVARVGLLVGLEGINITLTGTPVHQL	120
Db	61	LFIGAEIVAVHFSAEWFFVGTANTNTSYKAFSARVTVARVGLLVGLEGINITLTGTPVHQL	120

QY 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYA 180
Db 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYA 180
QY 181 SATLWVAFCEFWLNSVNLSTPAPLYGGALITTTGAFALFGVPALASISSVPLCPRLGSS 240
Db 181 SATLWVAFCEFWLNSVNLSTPAPLYGGALITTTGAFALFGVPALASISSVPLCPRLGSS 240
QY 241 ALTTQYGAFAFWTLATGVLCLFLGGAVVSLOQYVRPSALRTLDDQSAKDCSQERGSPLIL 300
Db 241 ALTTQYGAFAFWTLATGVLCLFLGGAVVSLOQYVRPSALRTLDDQSAKDCSQERGSPLIL 300
QY 301 GDPLHKQAAALPDLKCITTNL 320
Db 301 GDPLHKQAAALPDLKCITTNL 320

RESULT 2

US-10-264-237-2502
; Sequence 2502, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131PI
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2502
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (106)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-237-2502

Query Match 99.6%; Score 1660; DB 15; Length 320;
Best Local Similarity 99.7%; Pred. No. 6.3e-155;
Matches 319; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTLMNGVLPFYQPQPRHAGFSVPLILVILVFLAASFLILPGIRGSRMFWLVRLLS 60
Db 1 MTLMNGVLPFYQPQPRHAGFSVPLILVILVFLAASFLILPGIRGSRMFWLVRLLS 60
QY 61 LFIGAEIVAVHFSAEWFGVIVNTNTSYKAFSAARVTARVGLVGLGGINITLTGTPVHQ 120
Db 61 LFIGAEIVAVHFSAEWFGVIVNTNTSYKAFSAARVTARVGLVGLGGINITLTGTPVHQ 120
QY 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYA 180
Db 121 NETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYA 180
QY 181 SATLWVAFCEFWLNSVNLSTPAPLYGGALITTTGAFALFGVPALASISSVPLCPRLGSS 240
Db 181 SATLWVAFCEFWLNSVNLSTPAPLYGGALITTTGAFALFGVPALASISSVPLCPRLGSS 240
QY 241 ALTTQYGAFAFWTLATGVLCLFLGGAVVSLOQYVRPSALRTLDDQSAKDCSQERGSPLIL 300
Db 241 ALTTQYGAFAFWTLATGVLCLFLGGAVVSLOQYVRPSALRTLDDQSAKDCSQERGSPLIL 300
QY 301 GDPLHKQAAALPDLKCITTNL 320
Db 301 GDPLHKQAAALPDLKCITTNL 320

RESULT 3

US-09-759-130B-425
; Sequence 425, Application US/09759130B
; Publication No. US20030022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: Kirst, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodearl, Andrew
; APPLICANT: Holtzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: MP100-535OMNIM
; CURRENT APPLICATION NUMBER: US/09/759,130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 425
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-130B-425

Query Match 52.5%; Score 874; DB 10; Length 343;
Best Local Similarity 58.5%; Pred. No. 1.6e-77;
Matches 172; Conservative 40; Mismatches 74; Indels 8; Gaps 3;

QY 9 PFYQPRHAGFSVPLILVILVFLAASFLILPGIRGSRMFWLVRLLSLFIGAEIV 68
Db 9 PFYQPRHAGFSVPLILVILVFLAASFLILPGIRGSRMFWLVRLLSLFIGAEIV 68
QY 69 AVHPSAEWFGVIVNTNTSYKAFSAARVTARVGLVGLGGINITLTGTPVHQINETIDYNE 128
Db 69 AVHPSAEWFGVIVNTNTSYKAFSAARVTARVGLVGLGGINITLTGTPVHQINETIDYNE 128
QY 129 QFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYASATLWYAF 188
Db 129 QFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYASATLWYAF 188
QY 189 CFWLSNVNLSTPAPLYGGALITTTGAFALFGV--FALASISSVPLCPRLGSSALTTQY 246
Db 189 CFWLSNVNLSTPAPLYGGALITTTGAFALFGV--FALASISSVPLCPRLGSSALTTQY 246
QY 247 GAFAFWTLATGVLCLFLGGAVVSLOQYVRPSALRTLDDQSAK-----DCSQERGG 295
Db 247 GAFAFWTLATGVLCLFLGGAVVSLOQYVRPSALRTLDDQSAK-----DCSQERGG 295

PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: US 09/333,159
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: US 09/596,194
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 09/342,364
PRIOR FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: US 09/608,452
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/393,996
PRIOR FILING DATE: 1999-09-10
PRIOR APPLICATION NUMBER: US 09/602,871
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 09/420,707
PRIOR FILING DATE: 1999-10-19
NUMBER OF SEQ ID NOS: 460
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 427
LENGTH: 112
TYPE: PRT
ORGANISM: Homo sapiens
US-09-759-130B-427

Query Match 27.5%; Score 458; DB 10; Length 112;
Best Local Similarity 76.8%; Pred. No. 3.2e-37;
Matches 86; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

QY 71 HFSAEWFGVTNTNTSYKAFSAARVTARVGLLVGEGINTLTGTPVHQLNETIDYNEQF 130
DB 1 NFSSEWSVGQVSTNTSYKAFSSSEWISADIGLVGLGVNITLTGTPVQQLNETINYNBEF 60

QY 131 TWRLKENYAAEYANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYASA 182
DB 61 TWRLGENYAECAKALEKGLPDPVLYLAKEFTPRSPCGLYRQYRLAGHYTSA 112

RESULT 7

US-10-042-431-57
Sequence 57, Application US/10042431
Publication No. US20020182675A1
GENERAL INFORMATION:
APPLICANT: MCCARTHY, Sean A
APPLICANT: BARNES, Thomas M
APPLICANT: FRASER, Christopher C
APPLICANT: SHARP, John D
TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC,
TITLE OF INVENTION: PREVENTIVE, THERAPEUTIC, AND OTHER USES
FILE REFERENCE: 10147-6U2
CURRENT APPLICATION NUMBER: US/10/042,431
CURRENT FILING DATE: 2001-10-25
PRIOR APPLICATION NUMBER: US 09/333,159
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: US 09/578,063
PRIOR FILING DATE: 2000-05-24
NUMBER OF SEQ ID NOS: 79
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 57
LENGTH: 112
TYPE: PRT
ORGANISM: Homo sapiens
US-10-042-431-57

Query Match 27.5%; Score 458; DB 13; Length 112;
Best Local Similarity 76.8%; Pred. No. 3.2e-37;
Matches 86; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

QY 71 HFSAEWFGVTNTNTSYKAFSAARVTARVGLLVGEGINTLTGTPVHQLNETIDYNEQF 130
DB 1 NFSSEWSVGQVSTNTSYKAFSSSEWISADIGLVGLGVNITLTGTPVQQLNETINYNBEF 60

QY 131 TWRLKENYAAEYANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYASA 182
DB 61 TWRLGENYAECAKALEKGLPDPVLYLAKEFTPRSPCGLYRQYRLAGHYTSA 112

RESULT 8

US-09-759-130B-410
Sequence 410, Application US/09759130B
Publication No. US20030022279A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: McCarthy, Sean A
APPLICANT: Fraser, Christopher C
APPLICANT: Sharp, John D
APPLICANT: Barnes, Thomas S
APPLICANT: Kirst, Susan J
APPLICANT: Mackay, Charles R
APPLICANT: Myers, Paul S
APPLICANT: Leiby, Kevin R
APPLICANT: Wrighton, Nicolas
APPLICANT: Goodearl, Andrew
APPLICANT: Holtzman, Douglas A
TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
FILE REFERENCE: MP100-535OMNIM
CURRENT APPLICATION NUMBER: US/09/759,130B
CURRENT FILING DATE: 2002-09-16
PRIOR APPLICATION NUMBER: US 09/479,249
PRIOR FILING DATE: 2000-01-07
PRIOR APPLICATION NUMBER: US 09/559,497
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: US 09/578,063
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: US 09/333,159
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: US 09/596,194
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 09/342,364
PRIOR FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: US 09/608,452
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/393,996
PRIOR FILING DATE: 1999-09-10
PRIOR APPLICATION NUMBER: US 09/602,871
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 09/420,707
PRIOR FILING DATE: 1999-10-19
NUMBER OF SEQ ID NOS: 460
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 410
LENGTH: 383
TYPE: PRT
ORGANISM: Homo sapiens
US-09-759-130B-410

Query Match 16.6%; Score 276.5; DB 10; Length 383;
Best Local Similarity 29.9%; Pred. No. 1.2e-18;
Matches 83; Conservative 43; Mismatches 125; Indels 27; Gaps 8;

QY 27 VILVFLAALASFLILPGIRGSRWFMLVRVLLSLFIGAETVAVHFSAEWFGVTNTNTS 86
DB 14 IFSVFLIPLIAYITLIPGVR-RKRVTTVTYVLMVAVGALIASLTPCWSGSQMIYIQ 72

QY 87 YKAFSAARVTARVGLLVGEGINTLTGTPVHQLNETID-----YNEQFTWRLKENY 138
DB 73 FRGHSNERILAKIGVEIGLQKXAVTLKFERLLSSNDVLPGSDMTLYYNEGFDISGISM 132

QY 139 AAEYANALEKGLPDPVLYLAKEFT--PSSPCGLYHQYHLAGHYASATLWVAFCFWLISN 195
DB 133 AEAHLHGLNGLPYFPLSVLEYSLNQSFDWG--RHYRVAGHYTHAATWFAFACWCLSV 190

QY 196 VL-LSTPAPLYGGLALLTGAFALFGVFALASISSVPLCPRLGSSA-----LTTQYG 247
DB 191 VLMLPLPHNAY--KSILATGISCLIACLVYLLISP---CELRIFAFTGENPERVDLTATFS 245

```

Oy      248 AAFWWTATGVCLEFLGAGVSLQYVRPSAIRTLDDQS 2855
          |::|::|::|::|::|::|::|::|::|::|
Db      246 FCFYLLFAIGILCVLCGLGLGICEHHMRYITSTFLDAS 283

```

RESULT 9

```

US-10-042-431-40
; Sequence 40, Application US/10042431
; Publication No. US20020182675A1
; GENERAL INFORMATION:
; APPLICANT: MCCARTHY, Sean A
; APPLICANT: BARNES, Thomas M
; APPLICANT: FRASER, Christopher C
; APPLICANT: SHARP, John D
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC
; TITLE OF INVENTION: PREVENTIVE, THERAPEUTIC, AND OTHER USES
; FILE REFERENCE: 10147-6U2
; CURRENT APPLICATION NUMBER: US/10/042,431
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 40
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-042-431-40

```

Query Match	16.6%	Score 276.5;	DB 13;	Length 383;
Best Local Similarity	29.9%;	Pred. No. 1.2e-18;		
Matches	83;	Conservative	43;	Mismatches 125;
			Indels	27;
			Gaps	8;

[illegible]

RESULT 10

US-10-156-761-9442
 ; Sequence 9442, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, IKUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089

```

; PRIOR FILING DATE: 2001-05-30
;
; PRIOR APPLICATION NUMBER: JP 2001-272697
;
; PRIOR FILING DATE: 2001-08-02
;
; NUMBER OF SEQ ID NOS: 15109
;

```

Query Match	6.8%;	Score 113.5;	DB 14;	Length 792;
Best Local Similarity	24.2%;	Pred. No. 0.034;		
Matches 79;	Conservative 31;	Mismatches 111;	Indels 105;	Gaps 15;

QY	25	LIVILVFLALAA	SFLLILPGIRGSRKFWLVRVL	LSLFICAEIVAV-----HFSABEW	FV78
Dd	221	LDVLGALLLA	VASTCAVLLTSMGGTEYAMD	KRIILGLGAGACVA	AVLFLVEHFAP
QY	79	GTVNTN	TSYKAF--SARVTARVGLLVGLEGINITLTGT	PRHQNLNETIDYN	EQFTWR
Dd	281	-----PLRLFRDSVF	NVTGVLELVIG-----VALFG-----		
QY	137	NYAAEY-----ANALEKEJLDP	PVLYLAEXFTPSSPCGLYHQYHLA	GHYASAT---	1833
Dd	307	--AASYLP	TFLOQWVDGASATESGL---LML-----PMMAGI	VGASIIAGQLISHTGRY	354
QY	184	-LW-----VAFCEWLLSNV	LLSTPAPLYG-GLALLTTGA	FALFGVFALASISSV	PLCP234
Dd	355	RMPPIGSV	LAAVGMWLLSRLEADTPRLQYSIMWA	VLGAGIGMVPVLVLA	VQNVSVR--P412
QY	235	LRIGSSAL	TTQYGAAFVWTLATGVLCFLG	GAVVSLQYVRPEALRTILD	OSAKDCSOERG294
Dd	413	ADLGTA	TSAANNYFRQ-----IGSGVGA-----AVFGTL	FAHR	LADALBER-452
QY	295	GSPLLIGD	PLHKQAALPDLKCITTNL	320	
Dd	453	-----LPTRAGAR	LDPDESTLPOL471		

RESULT 11

```

US-09-815-242-13816
; Sequence 13816, Application US/09815242
; Patent No. US2002006156A1
;
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
;
; TITLE OF INVENTION: Identification of Essential Genes in
; Prokaryotes
;
; FILE REFERENCE: ELITRA.011A
;
; CURRENT APPLICATION NUMBER: US/09/815,242
;
; CURRENT FILING DATE: 2001-03-21
;
; PRIOR APPLICATION NUMBER: 60/191,078
;
; PRIOR FILING DATE: 2000-03-21
;
; PRIOR APPLICATION NUMBER: 60/206,848
;
; PRIOR FILING DATE: 2000-05-23
;
; PRIOR APPLICATION NUMBER: 60/207,727
;
; PRIOR FILING DATE: 2000-05-26
;
; PRIOR APPLICATION NUMBER: 60/242,578
;
; PRIOR FILING DATE: 2000-10-23
;
; PRIOR APPLICATION NUMBER: 60/253,625
;
; PRIOR FILING DATE: 2000-11-27
;
; PRIOR APPLICATION NUMBER: 60/257,931
;
; PRIOR FILING DATE: 2000-12-22
;
; PRIOR APPLICATION NUMBER: 60/269,308
;
; PRIOR FILING DATE: 2001-02-16
;
; NUMBER OF SEQ ID NOS: 14110
;
; SOFTWARE: FastSeq for Windows Version 4.0

```


SEQ ID NO 13816
LENGTH: 652
TYPE: PRT
ORGANISM: Salmonella typhi
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(652)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13816

Query Match 6.2%; Score 104; DB 9; Length 652;
Best Local Similarity 22.0%; Pred. No. 0.23;
Matches 63; Conservative 39; Mismatches 97; Indels 88; Gaps 15;

QY 65 AEIVAVH-----FSAEMVGTNTNTSYKAFSARVTARVGLVG-----LE 106
DB 153 ARIGAVHSAVIFGFSPEAIGRIIDSSRLV-----ITADEGVARGRSIPLKKNVDALK 207
QY 107 GINITLTGTPV--HQLNETIDYNE--QFTWRLKENYAAYANALEKGLP-----D 152
DB 208 NPNVTSVEHVIVLKRTGNDIDWQEGRLMWR-----DLIEKASPEHQPEAMNAED 257
QY 153 PVLVLAKEFTPPSSPCGLVHQYHLAGH--YASATLWVAF-----CFWLLSNVLLSTPAPL 204
DB 258 PLFILYTSGETGPKGVLT--TGGLVYAATTFKYVEDYHPGDIYCTADV----- 307
QY 205 YGGLALLTTGAFALFVGFALASISSVPLCPRLGSSALTTQYGAFFWTATGVLCLFLG 264
DB 308 ---GWVTHSHYLLYG-----FLACGATLMFE--GVENWPTPAR--MCQVVD 347
QY 265 GAVVSLQYVRPSALRTLDDQSAKDCSQERGSSPLL---GDPLHKQA 308
DB 348 KHQVNILYTABTALRALMMEGDKAIEGTDRSSLRLIGSVGEPIPEA 394

RESULT 12

US-10-156-761-8917
Sequence 8917, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 8917
LENGTH: 502
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-8917

Query Match 6.2%; Score 103.5; DB 14; Length 502;
Best Local Similarity 20.7%; Pred. No. 0.18;
Matches 73; Conservative 41; Mismatches 102; Indels 137; Gaps 15;

QY 6 GVLBFYQPRHAAG-----FSVLLVILVFLAALASFLILPGIRGHSRWFVLVRV 57
DB 188 GVLFAVPASRTKGTGRDGLGAFLLMLV-----LILLPIQGH--EWGWTSGR 235
QY 58 LLSLFIGAEIVAVHSAEWFGTNTNTSYK-----AFSARVTARVGLVGLGEGI 108
DB 236 TLGAFVGAIVMAV---VWVI-----TESRVKEPMVDMRFAHRPVLFTNLGALLGF--- 284

QY 109 NITLTGTPVHQLNETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPPSSPCG 168
DB 285 -----AMFTQFIGSVSLVQ-----MPEDV----- 303
QY 169 LYHQYHLAGHYASATLWVAFCFWLLSNVLLSTPAPLYGGL-----ALLTGAPALFGV- 221
DB 304 -----AGYFGASVLAASVYLLPTTLVSLVGAQFGVLVRLGARVTLAAGACFVL 356
QY 222 -----FALAS-----ISSVPL--CPRLGSSALTTQYG 247
DB 357 GFTWLSAHDATASVIGAGWITGLAISHGYASMPALIVASVPAHQGTIANGINSISRSVG 416
QY 248 AAFWVTLATGVLCLFLGAVVSLQYVRPSALRTLDDQSAKDCSQERGSSPLL 300
DB 417 SA---VASAVITSLASKTIPL---PDGMPALPQESQFTLSFTIAGAFAVL 461

RESULT 13

US-10-106-698-7554
Sequence 7554, Application US/10106698
Publication No. US20030109690A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypep
FILE REFERENCE: PA005P1
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 7554
LENGTH: 98
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (65)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (85)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (88)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC_FEATURE
LOCATION: (98)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-7554

Query Match 6.2%; Score 103; DB 14; Length 98;
Best Local Similarity 80.8%; Pred. No. 0.023;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 43 PGIRGHSRWFVLVRVLLSLFTGAEIV 68
DB 1 PGCLFLQRMFWLVRVLLSLFTGAEIV 26

RESULT 14

US-10-369-493-15557
Sequence 15557, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15557
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-15557

Query Match 6.2%; Score 103; DB 15; Length 460;
Best Local Similarity 21.7%; Pred. No. 0.18;
Matches 69; Conservative 39; Mismatches 106; Indels 104; Gaps 15;

QY 20 FSVPLILVILVFLAASFL-----ILPGIRGHSR--WF--WLVRVLLSL----- 61
DB 156 FSLIKVTVTVFIVLGGAAVGMPLADGSPAPGVR-HLRADGMFAHGTVPIIMTWAVN 214
QY 62 --FIGAEIVAVHFS-----AEMFVGT-----VNTNT 85
DB 215 FAFSGTELGIAAGETAQPARAIPLAIRTLRLVLFVGTVLVLAALLPAHAAAVETSP 274
QY 86 SYKAFSAARVTARVGLLVGLEGINITLTGTPVHQNETIDYNEQFTWRL--KENYAAEYA 143
DB 275 FVRAFELGIPYAAGLL-----NAVILTALISAANSGLYAAARMWLSLANEGTLPARFA 328
QY 144 NALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYASATLMVA----- 187
DB 329 RLTRRGIPLPALVL-----SMLGILLAL--LTGVYAADTVFVAISAVSGFAVVVWLSI 380
QY 188 ----FCF--WLSNVLLSTPA-----PLYGG-LALITTGAFALFGVFPALASISSV 230
DB 381 CASHYCFRRQLLRDGIALDSLAVRAPWYPWTPLIGALCVLACAGLAFDPQORIALMWCI 440
QY 231 PLCPRLGSSALTTQYGA 248
DB 441 PFVALCYGAHAVTQRLAA 458

RESULT 15
US-10-369-493-15930
; Sequence 15930, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15930
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-369-493-15930

Query Match 6.2%; Score 103; DB 15; Length 460;
Best Local Similarity 21.7%; Pred. No. 0.18;
Matches 69; Conservative 39; Mismatches 106; Indels 104; Gaps 15;

QY 20 FSVPLILVILVFLAASFL-----ILPGIRGHSR--WF--WLVRVLLSL----- 61

DB 156 FSLIKVTVTVFIVLGGAAVGMPLADGSPAPGVR-HLRADGMFAHGTVPIIMTWAVN 214
QY 62 --FIGAEIVAVHFS-----AEMFVGT-----VNTNT 85
DB 215 FAFSGTELGIAAGETAQPARAIPLAIRTLRLVLFVGTVLVLAALLPAHAAAVETSP 274
QY 86 SYKAFSAARVTARVGLLVGLEGINITLTGTPVHQNETIDYNEQFTWRL--KENYAAEYA 143
DB 275 FVRAFELGIPYAAGLL-----NAVILTALISAANSGLYAAARMWLSLANEGTLPARFA 328
QY 144 NALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYASATLMVA----- 187
DB 329 RLTRRGIPLPALVL-----SMLGILLAL--LTGVYAADTVFVAISAVSGFAVVVWLSI 380
QY 188 ----FCF--WLSNVLLSTPA-----PLYGG-LALITTGAFALFGVFPALASISSV 230
DB 381 CASHYCFRRQLLRDGIALDSLAVRAPWYPWTPLIGALCVLACAGLAFDPQORIALMWCI 440
QY 231 PLCPRLGSSALTTQYGA 248
DB 441 PFVALCYGAHAVTQRLAA 458

Search completed: February 23, 2004, 19:32:06
Job time : 627 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:17:08 ; Search time 45 Seconds
(without alignments)
684.029 Million cell updates/sec

Title:	US-09-936-456-2
Perfect score:	1666
Sequence:	1 MTLNNGVLPFYDQPRHAGF.....GDPLHQAALPDLKITTNL 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database :

```
PIR_78:*
1:  p1r1:*
2:  p1r2:*
3:  p1r3:*
4:  p1r4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	276.5	16.6	383	2	A88533	hypothetical prote
2	127.5	7.7	391	2	B68808	multidrug-efflux t
3	121.5	7.3	401	2	D83760	hypothetical prote
4	115	6.9	467	2	AG1059	probable membrane
5	115	6.9	815	2	T35970	probable efflux pr
6	113	6.8	509	2	T11043	cytochrome-c oxida
7	111.5	6.7	381	2	T11546	ubiquinol-cytochro
8	104	6.2	652	2	AB1020	acetate-CoA ligase
9	104	6.2	703	2	T13393	NADH2 dehydrogenas
10	103.5	6.2	456	1	S27616	probable glucarate
11	102.5	6.2	322	2	AF2066	hypothetical prote
12	102	6.1	358	2	G75586	urea/short chain-a
13	101.5	6.1	342	2	D84180	hypothetical prote
14	101.5	6.1	465	2	AG0648	nitrite extrusion
15	101	6.1	403	2	CG8422	nitrate transporte
16	101	6.1	546	2	AE0571	probable membrane
17	101	6.1	730	2	T44313	hypothetical prote
18	101	6.1	1074	2	T45094	probable arabinosy
19	100.5	6.0	381	2	T11312	ubiquinol-cytochro
20	100.5	6.0	414	2	F65097	hypothetical 43.5
21	100.5	6.0	414	2	C91125	probable transport
22	100.5	6.0	414	2	B85970	probable transport
23	100.5	6.0	414	2	AG0894	probable membrane
24	100	6.0	730	2	S70954	otng protein - Vib
25	99.5	6.0	398	2	G97744	cytochrome b limpo
26	99.5	6.0	422	2	AF0072	probable symporter
27	99.5	6.0	536	2	E82559	virulence factor X
28	99.5	6.0	958	2	AC0204	probable integral
29	99	5.9	612	2	G64678	NADH2 dehydrogenas

30	99	5.9	734	1	DERZN5	NADH2 dehydrogenas
31	98.5	5.9	378	1	S01190	ubiquinol-cytochro
32	98.5	5.9	379	1	S43267	ubiquinol-cytochro
33	98	5.9	381	1	A53224	ubiquinol-cytochro
34	98	5.9	449	2	S71005	glutamate transpor
35	98	5.9	463	1	GRECNK	nitrite extrusion
36	98	5.9	463	2	H90844	nitrite extrusion
37	98	5.9	463	2	G85702	nitrite extrusion
38	98	5.9	702	2	T13655	NADH2 dehydrogenas
39	97.5	5.9	353	2	T01542	hypothetical prote
40	97.5	5.9	406	2	T43120	conserved hypothet
41	97.5	5.9	432	2	S38893	uracil transport p
42	97.5	5.9	694	2	T13572	NADH2 dehydrogenas
43	97.5	5.9	704	2	T13503	NADH2 dehydrogenas
44	97.5	5.9	706	2	T13391	NADH2 dehydrogenas
45	97	5.8	393	1	CBPOM	ubiquinol-cytochro

ALIGNMENTS

RESULT 1

hypothetical protein C06E1.3 - *Caenorhabditis elegans*
C/Species: *Caenorhabditis elegans*
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 02-Nov-2001
C/Accession: A88533
R/anonymous: The *C. elegans* Sequencing Consortium.
Science 282, 2012-2018, 1998
A/Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biology
A/Reference number: A75000; MUID:99069613; PMID:9851916
A/Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A/Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A/Accession: A88533
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-383 <STO>
A/Cross-references: GB:chr_III; PIDN:AAA27934.1; PID:g289628; GSPDB:GN00021
A/Note: putative
C/Genetics:
A/Map position: 3

Query Match	16.6%	Score 276.5	DB 2	Length 383	
Best Local Similarity	29.9%	Pred. No. 2.2e-15			
Matches 83	Conservative 43	Mismatches 125	Indels 27	Gaps 8	

QY	27	VILVFIALAASFLLLPGIRGSRWFVLVRLLSLFGAIVAVHPSAEWFGVTNNTS	86
		: : : : : : : : : : :	
Db	14	IFSVPFLPIAYILPLGVR-RKRVTTVTVYVLMVAGALLIASLIYPCWASGQMIYIQ	72
QY	87	YKAFSARVTVARVGLVGLGGINITLTGTPVHQNETID-----YNEQFTWRLKENY	138
		: : : : : : : : : : : : : : : :	
Db	73	FRGHSNERILAKIGVEIGLQKNVTLTKFERLLSNDVLPGSDMTELYNEGFDISGISM	132
QY	139	AAEYANALEKGLPDPVLYLAEKFT--PSPCGLYHQYHLAGHYASATLWVAFCEWLLSN	195
		: : : : : : : : : : : : : : :	
Db	133	AEALHHGHENGLPYPMLSVLEPFLSNQDSFDWG--RHRYVAGHYTHAIIWFAFACCLSV	190
QY	196	VL-LSTPAPLYGGIALLLTTGAFAALFGVFALASISSVPLCPRLGSSA-----LTTQYG	247
		: : : : : : : : : : : : : : : :	
Db	191	VLMFLPEHNAY--KSILATGISCLIACTVYLLSP---CELRIFAFTGENFERVDLTATFS	245
QY	248	AAFWVTIATGVLCFLGGAVVSLOQYRPSALRTLLDQS	285
		: : : : : : : : : : : : : : :	
Db	246	PCFYLLFAIGILCVICGLGIGICEHWRIYTLSTFLDAS	283

RESULT 2

multidrug-efflux transporter homolog ykfK - *Bacillus subtilis*
C/Species: *Bacillus subtilis*
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C/Accession: B69808

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C/Accession: T35970
R/Seeger, K.U.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, January 1999
A/Reference number: Z21551
A/Accession: T35970
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-815 <SEE>
A/Cross-references: EMBL:AL035161; PIDN:CAA22731.1; GSPDB:GN00070; SCODEB:SC9C7.19
A/Experimental source: strain A3(2)
C/Genetics:
A/Gene: SCODEB:SC9C7.19

Query Match 6.9%; Score 115; DB 2; Length 815;
Best Local Similarity 24.2%; Pred. No. 0.13;
Matches 85; Conservative 25; Mismatches 107; Indels 134; Gaps 15;

QY 12 POPHAAAGSVPLLVILVFLAASFLILPGIRGSRMFWLVRVLLSLFIGAE----- 66
DB 226 PRPR-----VKPRLDILGALLAAASTCLVLTSMGTEYAMGSRVILGLAGAVATLL 280
QY 67 -IVAVHFAEWFVGTNTNTSYKAF--SAARVTARVGLVGLGINITLTGTPTVHQLNET 123
DB 281 FLVAEHFAPEPLI-----PLRFRDSTFNTALVGLVVG-----VALFG----- 319
QY 124 IDYNEQFTWRLKENYAAY-----ANALEKGLPDPVLYLAEKFTPS----- 164
DB 320 -----AASYLPTFLQWVDGASATESGL-----LMLPMGWGVGASI 355
QY 165 -----SPCGLYHQYHLAGHYASATLWVAFCEWLLSNVLLSTPAPLYG-GLALLTTGAFA 217
DB 356 ISGQLISRTGHRHPIILGALSIV-----GMWLLSRLDADTSRLHYSIMWAVLGAGIGL 410
QY 218 LFGVFALASISSVPLCPRLRGSSALTTOY-----GAFFWTLATGVLCFLGAVVS 269
DB 411 VMPVLVLAVGNSVR--PTDGTATSAANNFRQIGSVGAAVFGTLFAG----- 456
QY 270 LQYVRPSALRTLIDQSAKDCQERGSSPLILGDLHKQALPDLKCTTNL 320
DB 457 -----RLTDALADRIPEAG-----VGLPDAEAITPQL 484

RESULT 6

T11043
cytochrome-c oxidase (EC 1.9.3.1) chain I - Chlamydomonas eugametos mitochondrion
C/Species: mitochondrion Chlamydomonas eugametos
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-May-2000
C/Accession: T11043
R/Denovan-Wright, E.M.; Nedelcu, A.M.; Lee, R.W.
Plant Mol. Biol. 36, 285-295, 1998
A/Title: Complete sequence of the mitochondrial DNA of Chlamydomonas eugametos.
A/Reference number: Z17244; MUID:98145434; PMID:9484440
A/Accession: T11043
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-509 <DEN>
A/Cross-references: EMBL:AF008237; NID:G2865253; PID:G2865259; PIDN:AA39342.1
C/Genetics:
A/Genome: mitochondrion
A/Introns: 365/3
A/Note: cox1
C/Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C/Keywords: copper; electron transfer; heme; iron; magnesium; membrane-associated complex
F:10-254/Domain: cytochrome-c oxidase chain I homology <CO1>
F:62,374/Binding site: heme a iron (His) (axial ligands) #status predicted
F:237,286,287/Binding site: copper (His) #status predicted
F:237-241/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F:241/Binding site: oxygen (Tyr) #status predicted
F:364/Binding site: magnesium (His) (shared with chain II) #status predicted
F:372/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 6.8%; Score 113; DB 2; Length 509;

Best Local Similarity 21.2%; Pred. No. 0.11;
Matches 66; Conservative 38; Mismatches 118; Indels 90; Gaps 14;

QY 26 IVLVFLAALASFLILPGIRGSRW-----FWLVRVLLSLFIGA 65
DB 64 IIMLFFWMPALF-----GGFGNWLVPILIGAPDMAFRLNNISFWLNPSALGLILS 116
QY 66 EIVAVHFAEWFVGTNTNTSYKAFSAARVTARVGLV-----GLE-----GINITLTGT 115
DB 117 TWEQAGACTGW-----TAYPPLSIQSTGAVDLALISHLNGLSSISGINILVTTA 168
QY 116 PVHQLNETIDYNEQFTWRLKENYAAYANALEKGLPDPVLYLA-----EKFTP 163
DB 169 GMRAIGMKLSQMPFLFWSIA-----FTALVLVLAAPVLAALVMLLTDRLNTAYFCE 221
QY 164 SSPCGLY-HQYHLAGHYASATLWVAFCEWLLSNVL-LSTPAPLYGGLALL-TTGAFAALFG 220
DB 222 SGDLILYQHLFWPFGH-PEVYILILPAFGIVSHVISFSGKPIFGNMGMICAMGAISILG 280
QY 221 VFALAS-----ISSVPLCPRLGS--SALTTOYGAAFWT---LATGV 258
DB 281 FIWMAHMFVTGDLDTIAYFTSATMTIAVPTGMKIFSWLATIYGSLMTTPMFAVG 340
QY 259 LCLFLGAVVSL 270
DB 341 ICLFTLGCVTV 352

RESULT 7

T11546
ubiquinol-cytochrome-c reductase (EC 1.10.2.2) cytochrome b - spiny dogfish mitochondrion
C/Species: mitochondrion Squalus acanthias (spiny dogfish)
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Jun-2002
C/Accession: T11546
R/Rasmussen, A.S.; Arnason, U.
J. Mol. Evol. 48, 118-123, 1999
A/Title: Phylogenetic studies of complete mitochondrial DNA molecules place cartilag.
A/Reference number: Z17281; MUID:99091711; PMID:9873084
A/Accession: T11546
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-381 <RAS>
A/Cross-references: EMBL:Y18134; NID:G4186095; PIDN:CAA77061.1; PID:G4186108
C/Genetics:
A/Genome: mitochondrion
A/Genetic code: SSCI
C/Superfamily: cytochrome b; cytochrome b homology; cytochrome b6 homology; plastocyanin
C/Keywords: electron transfer; heme; iron; metalloprotein; mitochondrion; oxidative phosphorylation
F:12-340/Domain: cytochrome b homology <CYB>
F:12-210/Domain: cytochrome b6 homology <CB6>
F:222-340/Domain: plastocyanin-plastocyanin reductase 17K protein homology <17K>
F:84,183/Binding site: heme iron (His) (axial ligands) (low potential) #status predicted
F:98,197/Binding site: heme iron (His) (axial ligands) (high potential) #status predicted

Query Match 6.7%; Score 111.5; DB 2; Length 381;
Best Local Similarity 25.8%; Pred. No. 0.11;
Matches 70; Conservative 35; Mismatches 103; Indels 63; Gaps 14;

QY 27 VILVFLAALASFL-LILPGIRGSRW---FWLVRVLLSL---FTGAIVAVHFAEWF 76
DB 118 VILFLMLATAFVGVLVLP-----WGQMSFWGATVITNLISAPFYIGMLV-----QW 164
QY 77 FVGTVN---TNTSYKAFSAARVTARVGL-LVGLGINITLTGTPTVHQLNETID----- 125
DB 165 IMGGSIDNATLTRFFAFHFLPFLIVGLTLHLFLHFGSNNPM-GLNSDMDKISFHP 223
QY 126 ---YNEQFTWRLKENYAAYANALEKGLPDPVLYLAEKFTSSPCGLYHQYHLAGHYASA 182
DB 224 YFSYKDLLGFFLMTIILLALLALFLPNLGD-----AENFIPANP--LVTPPHIKPEW---- 273
QY 183 TLWVAFCEWLLSNVLLSTPAPLYGGLALLTTGAFAALFGVFALASISSVPLCPRLRGSSAL 242
DB 274 -----YFLFAVALIRSIPIKLGVAL-----LFSIFLIMLIPMLHNSKQSRNIFRP 320

```
Qy      243  TTQYGAFWVTLTATGVLCLFLGGAVNSLQYV 273
          |||:::|
Db      321  MTQF--LFWTLVNAIILTWIGGQVPEQPF 349
```

RESULT

AB1020

acetate-CoA ligase (EC 6.2.1.1) - *Salmonella enterica* subsp. *enterica* serovar Typhi (serovar Typhi)
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi
A:Note: this species has also been called *Salmonella typhi*
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AB1020
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nucleotide 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AB1020
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-652 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD09259.1; PID:G16505263; GSPDB:GN00176
C:Genetics:
A:Gene: acs
C:Superfamily: acetate-CoA ligase; acetate-CoA ligase homology
C:Keywords: acid-thiol ligase; coenzyme A

Query Match	6.2%;	Score 104;	DB 2;	Length 652;
Best Local Similarity	22.0%;	Pred. No. 0.84;		
Matches	63;	Conservative	39;	Mismatches 97;
				Indels 88;
				Gaps 15;

```

QY      65 AEIVAVH-----FSAEMFVGTVNINTSYKAFSARAVTARVELLV-----LE 106
      |||      |||      |||      |||      |||
Db      153 ARIGAVHSVIFGGFSPEALAGRILIDSSRLV-----ITADEGVRAGRSIPLKNNVDALK 207
      |||      |||      |||      |||      |||
QY      107 GINITLTGTPTV--HQINETIDYNE--QFTWRLKENYAAEYANALEKGLP-----D 152
      |      |      |      |      |      |
Db      208 NPNVTSVEHVIVLKRTGNDIDMDQEGRDLWNR-----DLIEKASPEHQPEAMNAD 257
      |||      |||      |||      |||      |||
QY      153 PVLYLEAKFTPSSPCGIYHQYHLACH--YASATIMVAF-----CFWLLSNVLLSTPAPL 204
      |:|      |:|      |:|      |:|      |:|      |:|
Db      258 PLFILYTSGSTGPKGVLHT--TGGYLVYAATPEKYVFDYHPGDIYWCADY----- 307
      |||      |||      |||      |||      |||
QY      205 YGGLALLTTGAFALFGVAFALASISSVPLCPLRLGSSALLTQYGAAFWVTLATGVLCPLG 264
      |||      |||      |||      |||      |||      |||
Db      308 ---GWTGHSYLLYG-----PLACGATITLMFE-GVPMWPTPAR--MCQVVD 347
      |||      |||      |||      |||      |||
QY      265 GAVVSLQYVRPSALRTLIDOSAKDCSQERGGSP.LIL---GDPLHXQA 308
      ||:|      ||:|      ||:|      ||:|      ||:|      ||:|
Db      348 KHQVNIYLTAPTALRALMAEGDKALRGTDRSSLRILGVSVEPIINPEA 394

```

RESULT 9

TI3393

NADH2 dehydrogenase (ubiquinone) (BC1.6.5.3) chain 5 - Lithachne pauciflora chloroplast
C/Species: chloroplast Lithachne pauciflora
C/Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 03-Jun-2002
C/Accession: T13393
R/Clark, L.G.; Zhang, W.; Wendel, J.F.
Syst. Bot. 20, 436-460, 1995
A/Title: A phylogeny of the grass family (Poaceae) based on ndhF-sequence data.
A/Reference number: Z17570
A/Accession: T13393
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-703 <CL>
A/Cross-references: EMBL:U21976; NID:g755846; PID:g755847; PIDN:AAA64697.1
A/Experimental source: leaf
C/Genetics:
A/Genome: chloroplast

A;Note: ndhf
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C;Keywords: chloroplast; membrane-associated complex; NAD; oxidoreductase

Query Match

Best Local Similarity 22.4%; Pred. No. 0.92;
Matches 68; Conservative 47; Mismatches 102; Indels 86; Gaps 18;

```
QY      23 PLLIVLVFLALAA$FLLI-----LPGIRGHSRMFWLVRV----L$LFICAEIVA$HFS    73
           ||:::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db     82 PLTSIMULITTTVGILVLISDNMTSHDEGYLRFVVYISFNTSMGLVSI$NLIQIYFF    141
                                           ::||::||::||::||::||::||::||
QY      74 AE-----WEVGTVNNTNSYKAFSAARVTARV---GLLVLEGINITLTGT$P$VH    118
           |||||::||::||::||::||::||::||::||::||::||::||
Db     142 WELVGMCSYLLIGFMWETRAIAASACQKAF----VTNRVGD$FILLGILGF-FWITG----    1922
                                           ::||::||::||::||::||::||::||
QY      119 QLNETHIDYNEQFTWRLEKNYA$E-----YANALEKG$LDPDV-LYLAEKFT    1622
           :::::|::||::||::||::||::||::||::||::||::||::||
Db     193 ----SLEFRDLF-KIANNWIPDNGINSLLTTLCAFLFLGAVAKSAQ$PLHIWLPDAME    246
                                           ::||::||::||::||::||::||::||
QY      163 PSSP-CGIYHQYHTLAGHYASATLWVAFCFWLLSNVLLSTPA$LYGGIALITGAFALE-    219
           :||:|::||::||::||::||::||::||::||::||::||::||
Db     247 GPTPI$ALI$H-----AATWVAAGIFLLABRLPFLVP$PLIMS$VSL-VGITTLFELG    296
                                           ::||::||::||::||::||::||::||
QY      220 GV$PALA-----SISSVPLCPRLGSSALLTQYGAAEW--VTLATGVLCFLG-G    265
           ||||::||::||::||::||::||::||::||::||::||::||
Db     297 ATLLAQ$RDIKRSLAYSTM$QLGYMMLALGIG$---YQALLPHLITHAYS$KALLFLGSG    352
                                           ::||::||::||::||::||::||::||
QY      266 AVV    268
           :::
Db     353 SII    355
```

RESULT 10

\$27616

probable glutarate transporter - *Pseudomonas putida*
C/Species: *Pseudomonas putida*
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C/Accession: S27616
R/Burlingame, R.P.; Maruya, A.; Ally, A.; Ally, D.; Backman, K.C.
submitted to the EMBL Data Library, June 1992
A/Description: Nucleotide sequences of hydroxyproline-specific alpha-ketoglutarate
A/Reference number: S27612
A/Accession: S27616
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-456 <BUR>
A/Cross-references: EMBL:M69160; NID:g151313; PIDN:AAA25867.1; PID:g151314
C/Superfamily: hexuronate transporter
C/Keywords: membrane protein

Query Match	6.2%;	Score 103.5;	DB 1;	Length 456;
Best Local Similarity	22.8%;	Pred. No. 0.62;		
Matches	80;	Conservative	35;	Mismatches 111;
				Indels 125;
				Gaps 18

```

QY      19 GFSVPLILVL-----VFLALASFLILLPGIRGHSRMFWLVRVLLSLFICAEIVAVHESA 74
      |:::|
Db     101 GISTAVVLLFLLRFWGLAEAPSF-----PG-----NARIV-----A 132
      |:::|

QY      75 EWFVGTVNTNTSYKAFSAAR--VYARVGLLVGEGINITLT-----GTPVQINETIDY 126
      |:::|
Db     133 SWF-PIKERGTASAIENSAQYFATRAVRALDGLRLRLARAVRHGRPGHCVLAHLVD 191
      |:::|

QY     127 NEQFTWRLKENYAAEYANALEKGLD-----PVL-VIAEKTPSSPCGLYHQYH 174
      |:::|
Db     192 GDLRAERSPAGYAAAEVRSSPHGGLVDLEDSDKDKDGPKMDYIRQLLTNRMMGTY--- 247
      |:::|

QY     175 LAGHYASATLWVAFCFWLLSNVLLSTPAPLY---GGLALLTGAF-----L 218
      |:::|
Db     248 -LGQFCINALTYFFLTWF-----PVLVGERGMTILKAGIILASLPAICGFLGVLT 296
      |:::|

QY     219 FGVF-----ALASISSVPL-CDRLRIGSSALTTQYGAAFWVTLATGVLCLF----- 262
      |:::|

```


Db 297 GGVISDTLLRGNLSVARKPIVCGMVLNSMVICNVYDADMMVVCMAAFPGKAIGA 356
QY 263 LGGAVVSLQYVRPSALRTLDDQSAKDCSQERGG-----SPLIIG 301
Db 357 LGWAVVS-----DTSPKQIAGLSGGLFNTFENLSSISTPIITIG 394

RESULT 11

AF2066
hypothetical protein al12084 [imported] - Nostoc sp. (strain PCC 7120)
C/Species: Nostoc sp. PCC 7120
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C/Accession: AF2066
R/Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A/Reference number: AB1807; MUID:21595285; PMID:11759840
A/Accession: AF2066
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-322 <KUR>
A/Cross-references: GB:BA000019; PIDN:BA073783.1; PID:g17131175; GSPDB:GN00179
A/Experimental source: strain PCC 7120
C/Genetics:
A/Gene: al12084

Query Match 6.2%; Score 102.5; DB 2; Length 322;
Best Local Similarity 24.1%; Pred. No. 0.5;
Matches 77; Conservative 43; Mismatches 116; Indels 83; Gaps 16;

QY 3 LMNGVLP-----FYQPRHAGFSVPLIVIL---VFLALASFLILPGIRGSRMF 52
Db 8 LMWVPLPLVFLGYYYRVPHPPLSVLWFLGALSGFLALGLEVF-----DSVAN 60
QY 53 WLVA---RVLLSLF-----IGAEIVAVHFSAEWFGVTN-TNTSYKAFSAAR 94
Db 61 WVLNWQRIQRSLFGIALRQIAIGPIEEGCKILAVVANYYFORRYHUTSSSIFLFTIAA 120
QY 95 V---TAR---VGLVGLIGINITLTGTVPVQNLNETIDVNEQFTWRLKENYAA-----EYA 143
Db 121 ALGFTAEENWILYHGTATILDRSIGTPVHAM----FSAPWGVALGKYFHATERLNHYR 175
QY 144 NALEKGLPDPVLY-----LAEKFTPSSPCGLYHQYHLAGHYASATLWVAFCFW-----LL 193
Db 176 NLIVRAGINSVIFHALVNLVSSAWRYSPPURF-----LSYGFPLLVWV---FWRMEQLL 227
QY 194 SNVLLSTPAFLYGG-----LALLTTGAFALFGVAFALASISSVPLCPRLGSS 240
Db 228 ROYTSQRPPLNLISGHTSLHRYWQGLVLFALMLGNAIFGLFLARVIS-PLSPAQIFSP 286
QY 241 ALTTQYGAFFWTLATGVL 259
Db 287 DFEWLMSRFLTNLAFGIL 305

RESULT 12

G75586
urea/short chain-amide ABC transporter, permease protein - Deinococcus radiodurans (stra
C/Species: Deinococcus radiodurans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C/Accession: G75586
R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.U.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A/Reference number: A75250; MUID:20036896; PMID:10567266
A/Accession: G75586
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-358 <WHI>

A/Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12457.1; PID:g64
A/Experimental source: strain R1
A/Genetics:
A/Gene: DRA0322
A/Map position: 2
C/Superfamily: 1-arabinose transport system permease arah

Query Match 6.1%; Score 102; DB 2; Length 358;
Best Local Similarity 28.2%; Pred. No. 0.63;
Matches 66; Conservative 26; Mismatches 98; Indels 44; Gaps 11;

QY 3 LMNGV--LPFYQPRHAGFSVPLIVILVFLALASFLILPGIRGSRMFWLVRVLS 60
Db 94 LYNQVEKLPWFQPFHSAFPALMVVALPGLAFLIGWLMFRRRITG-----VFVS 144
QY 61 LFIGAEIVAVHFSAEWFG---TVNTN--TSYK-----AFSARVTVARVGLL-V 103
Db 145 IITQALVLAFA--VFWLSGAQGLTSGTNGITDQTFGLNMRSDAFARGLVATVLLAL 201
QY 104 GLEGINITLTGTVPVQNLNETIDVNEQFTWRLKENYAAEYANALEKGLPDPVLYLAEKFTP 163
Db 202 SLAG-TAWLLRTPFGSLTLTAVRDNENRTRFLGYNPAPAFKVAFLVG---GVL----- 249
QY 164 SSPCGLYHQYHLAGHYASATLWVAFCEFWLLSNVLLSTPAFLYGGALLTTGAFA 217
Db 250 SGLSGALYTLHL-GRISPAFIGVAFSIELVWVVALGGRASLNGAAGLVGELA 302

RESULT 13

D84180
hypothetical protein Vhg0197c [imported] - Halobacterium sp. NRC-1
C/Species: Halobacterium sp. NRC-1
C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C/Accession: D84180
R/Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky,
Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ji
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A/Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.
A/Title: Genome sequence of Halobacterium species NRC-1.
A/Reference number: A84160; MUID:20504483; PMID:11016950
A/Accession: D84180
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-342 <STO>
A/Cross-references: GB:AE004437; NID:g10579844; PIDN:AA018808.1; GSPDB:GN00138
C/Genetics:
A/Gene: VNG0197C

Query Match 6.1%; Score 101.5; DB 2; Length 342;
Best Local Similarity 19.6%; Pred. No. 0.65;
Matches 63; Conservative 43; Mismatches 117; Indels 99; Gaps 14;

QY 22 VPLLIVILVFLALASFL--LIPGIRGSRMFWLVRVLSLFIGAEIVAVHFSAEWFFV 78
Db 41 IPLVVV-----AGALMLGWIVDQTA-----LMIWRLIDITVSRSAILVFAQTSFA 88
QY 79 GTV-----NTNTSYKAFSAARVTVARVGLVGLIGIN-----ITLTGTPV 117
Db 89 NNITPQGAQGEFPAALLTTNVSDADYETA-----LAIAAGDALNFVPSIGLALTGTAA 143
QY 118 HQLNETIDYNEQFT-----WRLKENYAAEYANALEKGL----- 150
Db 144 YAIWTVTSPRLRSVVGIVGAFVLIVGGLAVAGWRLLRTTIGAAVVRVATPALRWLARVLP 203
QY 151 -----PDPVLYLAEKFTPS-----SPGCLYHQYHLAGHYASATLWVAFCEFWLLSNVLS 199
Db 204 VPVPSPERIEAGIEFVAVSVERLAADRTGVLKVIGLA--TAALLCGAIAWVVF-IALG 259
QY 200 TPAPLYGGLALLTTGAFALFGVF--ALASISSVPLCPRLGSSALLTQYGAA----- 249
Db 260 SPIPIVPLFVVPVGTMAIGIPTPGGLGIEGVHVALLTALTSAAPLVAAVVIRIGG 319

QY 250 FWTLTATGVLCLFLGAVVSLQ 271
DB 320 FWTMTVVG-----GGSLAVLR 335

RESULT 14

AG0648
nitrite extrusion protein (nitrite facilitator) [imported] - Salmonella enterica subsp.
C/Species: Salmonella enterica subsp. enterica serovar Typh
A/Note: this species has also been called Salmonella typhi
C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C/Accession: AG0648
R/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaara, P.
Nature 413, 848-852, 2001
A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A/Reference number: AB0502; MUID:21534947; PMID:11677608
A/Accession: AG0648
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-465 <PAR>
A/Cross-references: GB:AL513382; PIDN:CAD08370.1; PID:gl6502415; GSPDB:GN00176
C/Genetics:
A/Gene: STY1287
C/Superfamily: nitrate transport protein narX

Query Match 6.1%; Score 101.5; DB 2; Length 465;
Best Local Similarity 27.7%; Pred. No. 0.93;
Matches 33; Conservative 18; Mismatches 45; Indels 23; Gaps 5;

QY 160 KFTPSSPCGIYHQYHLAGHYASATLWV-----AFCFWLL-SNVLLSTPAPLYGGLAL 211
DB 17 EWRPEDPAFWQQRGH---RVASRNLWTSVPCLLAFECVWMLFSAVAVNLPKVGFN----- 68
QY 212 TTGAFALFGVAFALASISVPLCLPLRGSSALTTQYGAAFWVTLATGVL---CLFLGAV 267
DB 69 ---FTTDQLFMLTALPSYSGALLRVPSFMVPLFGGRRTAFSTGILIVPCVWLGFAY 123

RESULT 15

C83422
nitrate transporter PA1783 [imported] - Pseudomonas aeruginosa (strain PA01)
C/Species: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: C83422
R/Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A/Reference number: AB2950; MUID:20437337; PMID:10984043
A/Accession: C83422
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-403 <STO>
A/Cross-references: GB:AE004604; GB:AE004091; NID:G9947762; PIDN:AAG05172.1; GSPDB:GN001
A/Experimental source: strain PA01
C/Genetics:
A/Gene: nsaA; PA1783

Query Match 6.1%; Score 101; DB 2; Length 403;
Best Local Similarity 23.4%; Pred. No. 0.87;
Matches 69; Conservative 41; Mismatches 91; Indels 94; Gaps 16;

QY 21 SVPLILVLFALASFLILPGIRGSRWFLVRVLSLFIG---AEIVAVHFSAEW 76
DB 78 SAGILAQVIVIAALAAWL---GVNSYAO-----ALLGLFLGAGASFAVALPLASQW 129
QY 77 F-----VGTNTNTSYKAFSAARVTARVGL--LVGLEGINITLT-----G 114
DB 130 YPPQHQKAMGAGAGNSGVIVLALFAPGLAAAFGMNNVGFALLPLVLTLLVFAVARN 189

QY 115 TPVHQLNETH-DY-----NEQFTWRLKENYAAEYANALEKGLPDPVLYLAEKFTBSSPCGL 169
DB 190 APQRPAPKMGDYLKALGDRDSWWMFFYSVTFGGFL--GL-----ASTLPGX 235
QY 170 YH-QYHL-----AGHYASATLWVAFCEFWLLSNVLLSTPAPLYGGLALLTGAFALFGVFL 224
DB 236 FHDQYGLNPVYAGYTTACVFAAG-----SLMRPLGALADRIIGIRSLMTYTL 284
QY 225 ASISSVPLCLPLRLGSSALTTQYGAAFWVTLATGVLCLFL-----GGAVVSL 270
DB 285 AAI-----CIAAVG-----FHLPSAMAAALGLFVVAMLSLGANGAVFQL 323

Search completed: February 23, 2004, 19:20:31
Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 18:17:55 / Search time 48 Seconds

(without alignments)
347.134 Million cell updates/sec

Title: US-09-936-456-2

Perfect score: 1666
Sequence: 1 MTLNMGVLPFYPPQPRHAGF.....GDP LKQAALPDLCITTNL 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	286.5	17.2	397	YKQ3_CAEEL	P34298 caenorhabdi
2	111.5	6.7	381	CYB_SQUAC	Q92242 squalus aca
3	104	6.2	652	ACSA_SALTI	Q92110 salmonella
4	103.5	6.2	456	GUDD_PSEPU	P42205 pseudomonas
5	103	6.2	652	ACSA_SALTY	Q92Kf6 salmonella
6	102.5	6.2	380	CYB_RANCA	P16674 rana catesb
7	102.5	6.2	381	CYB_HETFR	P34869 heterodontu
8	101	6.1	1074	EMBC_MYCSM	Q50393 mycobacteri
9	100.5	6.0	288	CLN8_MOUSE	Q9quk3 mus musculi
10	100.5	6.0	381	CYB_CARPL	P34866 carcharhinu
11	100.5	6.0	381	CYB_SCYCA	O79413 scyliorhinu
12	100.5	6.0	414	YGJU_ECOLI	P42602 escherichia
13	99.5	6.0	380	CYB_RANNT	Q9t6r9 rana nigrom
14	99	5.9	734	NU5C_ORYSA	P12129 oryza sativ
15	98.5	5.9	378	CYB_DROME	P18935 drosophila
16	98.5	5.9	379	CYB_CAPMR	P1287 caperea mar
17	98.5	5.9	381	CYB_CARPO	P34867 carcharhinu
18	98	5.9	463	NMRK_ECOLI	P10903 escherichia
19	98	5.9	702	NU5C_POAPR	Q32880 poa pratens
20	97.5	5.9	380	CYB_RANDY	Q9t6r5 rana dybows
21	97.5	5.9	380	CYB_RANPL	Q9t6r8 rana plancy
22	97.5	5.9	432	PYRF_BACCL	P41006 bacillus ca
23	97	5.8	652	ACSA_ECO57	Q8x5t5 escherichia
24	96.5	5.8	378	CYB_COLOH	Q9mfns cocilliomya
25	96.5	5.8	379	CYB_IGUG	Q956t5 iguana igua
26	96.5	5.8	381	CYB_SPHLE	P34874 sphyrna lew
27	96.5	5.8	383	CYB_CAICR	Q9b205 calman croc
28	96.5	5.8	384	CYB_HYSAP	Q04910 hystrix afr
29	96.5	5.8	471	NORM_CAUCR	P58163 caulobacter
30	96.5	5.8	702	NU5C_SORBI	Q33066 sorghum bic
31	96	5.8	381	CYB_CHLRE	P23662 chlamydomon
32	95.5	5.7	326	NUIM_ASTPE	P23650 asterina pe
33	95.5	5.7	381	CYB_SPHRT	P34876 sphyrna tib

34	95.5	5.7	381	1	CYB_SPHTV	P34875 sphyrna tib
35	95.5	5.7	1149	1	ABAI_BOVIN	Q29449 bos taurus
36	95	5.7	359	1	LICD_SYNY3	Q08871 synechocyst
37	95	5.7	393	1	CYB_ARATH	P42792 arabidopsis
38	95	5.7	419	1	CYB_RHOVI	P81378 rhodospseudo
39	95	5.7	739	1	NU5C_WHEAT	Q95h46 triticum ae
40	94.5	5.7	378	1	CYB_ANOQU	P33501 anopheles q
41	94.5	5.7	378	1	CYB_MAMPR	P92658 mammutus p
42	94.5	5.7	380	1	CYB_RANAM	Q9t6r2 rana amuren
43	94.5	5.7	381	1	CYB_PRIGL	P34873 prionace gl
44	94.5	5.7	397	1	CYB_ORYSA	P14833 oryza sativ
45	94.5	5.7	530	1	UD17_HUMAN	Q9haw7 homo sapien

ALIGNMENTS

RESULT 1	ID	YKQ3_CAEEL	STANDARD	PRT	397 AA.
AC	P34258:				
DT	01-FEB-1994 (Rel. 28, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Hypothetical protein C06E1.3 in chromosome III.				
GN	C06E1.3.				
OS	Caenorhabditis elegans.				
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;				
OC	Rhabditidae; Peloderinae; Caenorhabditis.				
OX	NCBI_TaxID=6239;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Bristol N2;				
RX	MEDLINE=94150718; PubMed=7906398;				
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,				
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,				
RA	Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,				
RA	Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,				
RA	Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,				
RA	Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,				
RA	Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,				
RA	Sims M., Smailson N., Smith A., Smith M., Sonhammer E., Staden R.,				
RA	Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,				
RA	Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,				
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.				
RT	elegans.";				
RL	Nature 368:32-38(1994).				
RN	[2]				
RP	REVISIONS.				
RA	Waterston R.;				
RL	Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; L16559; AAA27934.2; -.				
DR	WormPep; C06E1.3; CE30481.				
KW	Hypothetical protein; Transmembrane.				
FT	TRANSMEM 26				
FT	TRANSMEM 57				
FT	TRANSMEM 191				
FT	TRANSMEM 218				
FT	TRANSMEM 261				
FT	TRANSMEM 281				
FT	TRANSMEM 397				
FT	SEQUENCE 397 AA; 43801 MW; 3E6A11E2A8264A56 CRC64;				
SQ					
Query Match	17.2%;	Score 286.5;	DB 1;	Length 397;	
Best Local Similarity	29.4%;	Pred. No. 7.5e-16;			

```

Matches 89; Conservative 47; Mismatches 138; Indels 29; Gaps 9;
QY 4 WNGVLPFFYQDPRHAA--GSSVPLLVILVFLAALASFLILPGIRGHSRWFWLVRLSL 61
DB 3 WFGGNPSPSDYENAAIPENMHAFFVPSVFLIPLIAYILILPGVR-RKRVTITVTVVLM 61
QY 62 FIGAIVAVHSAEMFVGVNTNTSYKAFSAARVAVGLVLEGINITLTGTPVHQLN 121
DB 62 AVGALIASLIYPCWASGSOMITQTGRHSNERILAKIGVEIGLQKVNVTIKPERLLSN 121
QY 122 ETID-----YNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFT---PSSPCGLY 170
DB 122 DVLPGSDMTELYNNEGFDSIGISSMAEALHGLENGLPYPMLSVLEFSLNQDSFDWG-- 179
QY 171 HQYHLAGHYASATLWVAFCFWLLSNVL-LSTPAPLYGGLALLTGAFALFGVFAALASISS 229
DB 180 RHYRVAGHYTHAIIWFAFACWCLSVLMFLPHNAV--KSLIATGISCLIACLVYLLSP 237
QY 230 VPLCPRLGSSA-----LTTQYGAFAFWTLATGVLCLFLGAVVSLQYVRPSALRTL 282
DB 238 ---CELRIAFTEGNERVDLTATFSFCFYLIFALGILCVLCLGIGICEHWRIVTLSTFL 294
QY 283 DQS 285
DB 295 DAS 297

```

RESULT 2

CYB_SQUAC
ID CYB_SQUAC STANDARD; PRT; 381 AA.

AC Q9Z242;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Cytochrome b.

GN MTCYB OR COB OR CYTB.

OS Squalus acanthias (Spiny dogfish).

OG Mitochondrion.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;

OC Elasmobranchii; Squala; Squalidae; Squalus.

OX NCBI_TaxID=7797;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99091711; PubMed=9873084;

RA Rasmussen A.S., Arnason U.;

RT "Phylogenetic studies of complete mitochondrial DNA molecules place

RT cartilaginous fishes within the tree of bony fishes.";

RL J. Mol. Evol. 48:118-123(1999).

CC -!- FUNCTION: Component of the ubiquinol-cytochrome c reductase

CC complex (complex III or cytochrome b-c1 complex), which is a

CC respiratory chain that generates an electrochemical potential

CC coupled to ATP synthesis (By similarity).

CC -!- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or

CC b562) is low-potential and absorbs at about 562, and heme 2 (or BH

CC or b566) is high-potential and absorbs at about 566 (By

CC similarity).

CC -!- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,

CC cytochrome c1 and the Rieske protein (By similarity).

CC -!- SIMILARITY: Belongs to the cytochrome b family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; Y18134; CAA77061.1; -.

DR PIR; T11546; T11546.

DR InterPro; IPR005798; Cytb_b6_C.

DR InterPro; IPR005797; Cytb_b6_N.

RT Pfam; PF00032; cytochrome_b_C; 1.

```

DR Pfam; PF00032; cytochrome_b_N; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE; PS00193; CYTOCHROME_B_QO; 1.
KM Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KN Heme.
FT METAL 84 84 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 183 183 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 197 197 IRON 2 (HEME B566 AXIAL LIGAND).
SQ SEQUENCE 381 AA; 43260 MW; BD251559304A7730 CRC64;

```

Query Match

Best Local Similarity 6.7%; Score 111.5; DB 1; Length 381;

Matches 70; Conservative 35; Mismatches 103; Indels 63; Gaps 14;

```

QY 27 VILVFLAASFL-LILPGIRGHSRW----FWLVRLISL-----FIGAIVAVHSAEM 76
DB 118 VILVFLMATAFVGYLP-----WGQNSFWGATVITNLLSAFPYIGDMLV-----QW 164
QY 77 FVGTVN---TNTSYKAFSAARVTARVGL-LVGLIGINITLTGTPVHQLNETID----- 125
DB 165 IWGQPSIDNATLTRFFAFHPLPLIVGLTLIHLFLHETGSNNPM-GLNSDMDKTSFHP 223
QY 126 ---YNEQFTWRLKENYAAYANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYASA 182
DB 224 YFSYKDLGLGFLLMILLALLALFLPNLGD-----AENFIIPANP--LVTTPHIKPEW--- 273
QY 183 TLMVAFCFWLLSNVLSTPAPLYGGLALLTGAFALFGVFAALASISSVPLCPRLGSSAL 242
DB 274 -----YFLFAVAILRSIPNKLGVALL-----LFSIFILMLIPMLHTSKORSNIIFRP 320
QY 243 TTYGAFAFWTLATGVLCLFLGAVVSLQYV 273
DB 321 MTQF--LFWTLVANAIIITWTIGQVPEQPI 349

```

RESULT 3

ACSA_SALTI
ID ACSA_SALTI STANDARD; PRT; 652 AA.

AC O821R0;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Acetyl-coenzyme A synthetase (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-

DE activating enzyme).

GN ACS OR STY4473 OR T4181.

OS Salmonella typhi.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Salmonella.

OX NCBI_TaxID=601;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CT18;

RX MEDLINE=21534947; PubMed=11677608;

RA Churchill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,

RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,

RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,

RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,

RA Felkwell T., Hamlin N., Hague A., Hien T.T., Holtroyd S., Jagels K.,

RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,

RA Ouail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,

RA Whitehead S., Barrell B.G.;

RT "Complete genome sequence of a multiple drug resistant Salmonella

RT enterica serovar Typhi CT18.";

RL Nature 413:848-852(2001).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Ty2 / ATCC 700931;

RX MEDLINE=22531367; PubMed=12644504;

RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,

RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;

RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2

RT and CT18.";

```
RL Bacteriol. 185:2330-2337(2003).
CC -1- FUNCTION: Enables the cell to use acetate during aerobic growth to
CC generate energy via the TCA cycle, and biosynthetic compounds via
CC the glyoxylate shunt. Acetylates cheY, the response regulator
CC involved in flagellar movement and chemotaxis (By similarity).
CC -1- CATALYTIC ACTIVITY: ATP + acetate + CoA = AMP + diphosphate +
CC acetyl-CoA.
CC -1- PTM: Acetylated. Deacetylation by the SIR2-homolog deacetylase
CC activates the enzyme (By similarity).
CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL627282; CAD09259.1; -
DR EMBL; AE016848; AAC071645.1; -
DR HAMAP; MF_01123; -; 1.
DR InterPro; IPR000873; AMP-bind.
DR Pfam; PF00501; AMP-binding; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00455; AMP BINDING; 1.
KM Ligase; Acetylation; Complete proteome.
FT ACT_SITE 517 517 BY SIMILARITY.
FT MOD_RES 609 609 ACETYLATION (BY SIMILARITY).
SQ SEQUENCE 652 AA; 72163 MW; 1586F56B2B1695DA CRC64;

Query Match 6.2%; Score 104; DB 1; Length 652;
Best Local Similarity 22.0%; Pred. No. 0.67;
Matches 63; Conservative 39; Mismatches 97; Indels 88; Gaps 15;

QY AEIVAVH-----FSAEWFVGTNTNTSYKAFSARVTVAGLVG-----LE 106
Db 153 ARIGAVHSVIFGGRSPFAIAGRIDDSSRLV-----ITADGVVRAGRSIPLKKNVDALK 207
QY 107 GINTLTGTPTV-HQJNETIDYNE--QFTWRLEKNYAAYANALEKLP-----D 152
Db 208 NPNTSVSEHVIIVLKRGTNDIDWQEGRDLMWR-----DLIEKASPEHQPEANNAED 257
QY 153 PVLVLAKEFTPSPCGLYHQYLAGH--YASATLWVAF-----CFWLSNVLLSTPAPL 204
Db 258 PLFLYLTSGTGKPKGVLTHT--TGGLVLYAATFEKYVFDYHPGDIYWCADY----- 307
QY 205 YGGLALLTTGAPALFGVFALASISSVPLCPRLGSSALTTQYGAFFWTATGVLGFLG 264
Db 308 ---GWVTGHSYLLYG-----PLACGATTLMEF-GVPMWPTPAR--MCQYVD 347
QY 265 GAVVSLQYVRPSALRTLLDQSAKDCSQERGGSPPLI--GDPLHKQA 308
Db 348 KHQVNIILYTAPTAIRALMAEGDKAIEGTRSSLRILGSGVEPPIPEA 394

RESULT 4
GUDP_PSEPU STANDARD; PRT; 456 AA.
AC P42205;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable glucarate transporter (D-glucarate permease).
GN GUDP.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PP3;
```

```
RA Burlingame R.P., Lauer G.D., Platz J.G., Rudd E.A., Ally A.,
RA Ally D., Backman K.C.;
RT "Nucleotide sequence of genes for glucarate dehydratase and
RT 5-keto-4-deoxyglucarate dehydratase from Pseudomonas putida pp3.";
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: UPTAKE OF D-GLUCARATE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: Belongs to the phthalate permease family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M69160; AAA25867.1; -.
DR PIR; S27616; S27616.
DR InterPro; IPR004744; Dgal_transporter.
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub_transporter.
DR Pfam; PF00083; sugar_tr_1.
DR TIGRFAMs; TIGR00893; 2A0114; 1.
DR PROSITE; PS00850; MFS; 1.
KM Transmembrane; Transport.
FT TRANSMEM 11 31 POTENTIAL.
FT TRANSMEM 42 62 POTENTIAL.
FT TRANSMEM 92 112 POTENTIAL.
FT TRANSMEM 246 266 POTENTIAL.
FT TRANSMEM 280 300 POTENTIAL.
FT TRANSMEM 317 337 POTENTIAL.
FT TRANSMEM 342 362 POTENTIAL.
FT TRANSMEM 381 401 POTENTIAL.
FT TRANSMEM 408 428 POTENTIAL.
SQ SEQUENCE 456 AA; 49779 MW; B5B80BB6608C581D CRC64;

Query Match 6.2%; Score 103.5; DB 1; Length 456;
Best Local Similarity 22.8%; Pred. No. 0.49;
Matches 80; Conservative 35; Mismatches 111; Indels 125; Gaps 18;

QY 19 GFSVPLILVIL---VFLAASFLILIPGIRGSRWFWLVLLSLFGAETVAVHPSA 74
Db 101 GISTAVVILFLRFVWGLAEAPSF---PG-----NARIV-----A 132
QY 75 EMFVGTVNTNTSYKAFSAAR--VTARVGLLVGEGINITLT-----GTPVHQLNETIDY 126
Db 133 SWP-PTXERGTSALITNSAQYFATRAVRALDGLDLHLRLAARVHRHGRPGCVLAHLVD 191
QY 127 NEQFTWRLEKNYAAYANALEKLPD-----PVL-YLAKEFTPSPCGLYHQYH 174
Db 192 GDLRAERSPAGYAABVRSPPHGLVDLEDSDKDKDGGPKMDYTRQLLTNRMMGITY---- 247
QY 175 LAGHYASATLWVAFCFWLSNVLLSTPAPLY-----GGLALLTGAF-----L 218
Db 248 -LGQFCINALTYFFLTWF-----PVLVQERGMTILKAGIIASLPALCGFLGVL 296
QY 219 FGVF-----ALASISSVPL-CPLRLGSSALTTQYGAFFWTATGVLCLF----- 262
Db 297 GGVISDTLLRRGNSLSVARKTPIVCGMVLMSMILICNYVDADMVVCFMALAFGKAIGA 356
QY 263 LGGAVVSLQYVRPSALRTLLDQSAKDCSQERGG-----SPILIG 301
Db 357 LGWAVVS-----DTSRKQIAGLSGGLFNTFGNLISISTPIIIG 394

RESULT 5
ACSA_SALTY STANDARD; PRT; 652 AA.
AC Q82KF6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
```

DE Acetyl-coenzyme A synthetase (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-
activating enzyme).
GN ACS OR STM4275.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
CX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
EX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
RN [2]
RP ACETYLATION BY SIR2 HOMOLOG, AND MASS SPECTROMETRY.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=22382096; PubMed=12493915;
RA Starai V.J., Celic I., Cole R.N., Boeke J.D., Escalante-Semerena J.C.;
RT "Sir2-dependent activation of acetyl-CoA synthetase by deacetylation
of active lysine.";
RL Science 298:2390-2392(2002).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS) AND OF ACETYLATED PROTEIN (2.3
ANGSTROMS).
RX MEDLINE=22515404; PubMed=12627952;
RA Gulick A.M., Starai V.J., Horowitz A.R., Homick K.M.,
Escalante-Semerena J.C.;
RT "The 1.75 A crystal structure of acetyl-CoA synthetase bound to
adenosine-5'-propylphosphate and coenzyme A.";
RL Biochemistry 42:2866-2873(2003).
CC -!- FUNCTION: Enables the cell to use acetate during aerobic growth to
generate energy via the TCA cycle, and biosynthetic compounds via
the glyoxylate shunt. Acetylates cheY, the response regulator
involved in flagellar movement and chemotaxis (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + acetate + CoA = AMP + diphosphate +
acetyl-CoA.
CC -!- SUBUNIT: Monomer.
CC -!- PTM: Acetylated. Deacetylation by the SIR2-homolog deacetylase
activates the enzyme.
CC -!- MASS SPECTROMETRY: MW=733.4; METHOD=MALDI; RANGE=607-612.
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE008900; AAL23099.1; -
DR PDB; 1PG3; 03-JUN-03.
DR PDB; 1PG4; 03-JUN-03.
DR StvGene; SG7777; acs.
DR HAMAP; MF_01123; -; 1.
DR InterPro; IPR000873; AMP-bind.
DR Pfam; PF00501; AMP-binding; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00455; AMP BINDING; 1.
KW Ligase; Acetylation; 3D-structure; Complete proteome.
FT ACT_SITE 517 517
FT MOD_RES 609 609 ACETYLATION.
SQ SEQUENCE 652 AA; 72152 MW; 347209D1D3D349D8 CRC64;

Query Match 6.2%; Score 103; DB 1; Length 652;
Best Local Similarity 22.0%; Pred. No. 0.81;

Matches 63; Conservative 39; Mismatches 97; Indels 88; Gaps 15;
QY AEIVAVH-----FSAENVGTVNTNTSYKAFSAARVTVRGLVG-----LE 106
DB 153 ARIGAVHSVIFGFGSPFAVAGRIIDSSRLV-----ITADEGVRAGRSIPLKAVDALK 207
QY 107 GINITLTGTPV--HQLNETIDYNE--QFTWRKENVAAEYANALEKLP-----D 152
DB 208 NENVTSVEHVIVLKRTGSDIDWQEGRDLMWR-----DLIEKASPEHQPEANVAD 257
QY 153 PVLVLAKEFTPSPCGLVHYGYLHAGH--YASATLWVAF-----CEWLSNVLSSTPAPL 204
DB 258 PFLILYTSSTGSKPKGVLRH--TGVLVYAATFKYVFDYHPGDIWCTADV----- 307
QY 205 YGGLALLTGAFALFGVFALASISSVPLCPRLGSSALTTQYGAAPVTLATGVLCFLG 264
DB 308 ---GWTGHSYLLYG-----PLACGATLME--GVPNWPPTPAR--MCQVVD 347
QY 265 GAVVSLQYVRPSALRTILDQSAKDCSQERGSPLIL---GDPIHKOA 308
DB 348 KHQVNIYLTAPTALFALMAEGDKAIEGTRSSLRLISGVGEPIINPEA 394
RESULT 6
CYB_RANCA
ID CYB_RANCA STANDARD; PRT; 380 AA.
AC P16674; Q9TER7;
DT 01-AUG-1990 (Rel. 15, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome b.
GN MTCYB OR COB OR CYTB.
OS Rana catesbeiana (Bull frog).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.
OX NCBI_TaxID=8400;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee J.-B., Lee H.-Y., Yang S.-Y.;
RT "Genetic relationships of mitochondrial cytochrome b gene among six
RT Korean Rana species.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 267-380 FROM N.A.
RX MEDLINE=88033615; PubMed=2444617;
RA Yoneyama Y.;
RT "The nucleotide sequences of the heavy and light strand replication
RT origins of the Rana catesbeiana mitochondrial genome.";
RL Nippon Ika Daigaku Zasshi 54:429-440(1987).
CC -!- FUNCTION: Component of the ubiquinol-cytochrome c reductase
CC complex (complex III or cytochrome b-c1 complex), which is a
CC respiratory chain that generates an electrochemical potential
CC coupled to ATP synthesis (By similarity).
CC -!- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or
CC b562) is low-potential and absorbs at about 562, and heme 2 (or BH
CC or b566) is high-potential and absorbs at about 566 (By
CC similarity).
CC -!- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
CC cytochrome c1 and the Rieske protein (By similarity).
CC -!- SIMILARITY: Belongs to the cytochrome b family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF205089; AAF17088.1; -
DR EMBL; D12694; BAA02191.1; -
DR InterPro; IPR005798; Cytb_b6_c.

RA Telenti A., Philipp W., Sreevatsan S., Bernasconi C., Stockbauer K.E.,
 RA Wieles B., Musser J.M., Jacobs W.R.;
 RT "The emb operon, a gene cluster of Mycobacterium tuberculosis involved
 RT in resistance to ethambutol.";
 RL Nat. Med. 3:567-570(1997).
 CC -|- FUNCTION: Arabinosyl transferase responsible for the
 CC polymerization of arabinose into the arabinan of arabinogalactan.
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -|- MISCELLANEOUS: This is one of the target of the anti-tuberculosis
 CC drug ethambutol [(S,S')-2,2'-(ethylenedimino)di-1-butanol; EMB].
 CC EMB is a first-line drug used to treat tuberculosis. EMB inhibits
 CC the transfer of arabinogalactan into the cell wall.
 CC -|- SIMILARITY: BELONGS TO THE EMB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; U46844; AAC45271.1; -.
 DR PIR; T45094; T45094.
 DR InterPro; IPR007680; Arab transf.
 DR Pfam; PF04602; arab_transf; 1.
 KW transiferase; Glycosyltransferase; Transmembrane; Cell wall;
 KM Antibiotic resistance.
 FT TRANSMEM 15 37 POTENTIAL.
 FT TRANSMEM 214 236 POTENTIAL.
 FT TRANSMEM 251 273 POTENTIAL.
 FT TRANSMEM 415 437 POTENTIAL.
 FT TRANSMEM 452 474 POTENTIAL.
 FT TRANSMEM 516 538 POTENTIAL.
 FT TRANSMEM 573 595 POTENTIAL.
 FT TRANSMEM 608 630 POTENTIAL.
 FT TRANSMEM 645 667 POTENTIAL.
 FT TRANSMEM 684 706 POTENTIAL.
 SQ SEQUENCE 1074 AA; 115391 MW; 68DB19FE8839C301 CRC64;

Query Match 6.1%; Score 101; DB 1; Length 1074;
 Best Local Similarity 20.1%; Pred. No. 2.1;
 Matches 62; Conservative 35; Mismatches 81; Indels 130; Gaps 14;

QY 27 VILVFLAALASFLILPGIRGHSRMF---WLVRVLSLFIGAEIVAVHFSAEHFGVTANT 83
 DB 224 VAMTVIALGALHVLDCADGRKHKRFLPSKWSMTPLDGLVSAMLVMMH-----FVGANTA 278
 QY 84 NTSYKAFSAARVTARVGLVGLGINTLTGTPTVQNLNETIDVNEQFTWRLKENYAAEYA 143
 DB 279 DDGY-ILTWARVSEHAGYMA-----NYRWMFG 304
 QY 144 NALEKGLPDPVLYIAEKTPSSPCGLYHQ-YHLAGHYASATLWVAF-----CFWLLS 194
 DB 305 -----TPSPFGWYDLDLALMAHVSTASVMKRPFTLLMGLACWVVIS 346
 QY 195 -----NVLSTPAPLYGGL-----ALLT----- 212
 DB 347 REVIPRLGAAGKASRAAMTAAGLFLAFLWPLINNGLRPEPIALGILLTWCSEVERGVATS 406
 QY 213 -----TGAFALF-GVFALASISS--VPLCPRLGSSALTTQYGAFFVTL---A 255
 DB 407 RLLPVAVALIIGALTTFSGPTGIAAVGALLVAIGPKTIVAAHVSRFG--YVALLAPIAA 464
 QY 256 TGVLCFL 263
 DB 465 AGTVTIFL 472

RESULT 9
 CLN8 MOUSE
 ID CLN8 MOUSE STANDARD; PRT; 288 AA.
 AC Q9QURK3;

DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE CLN8 protein.
 GN CLN8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99438402; Pubmed=10508524;
 RA Ranta S., Zhang Y., Ross B., Lonka L., Takkunen E., Messer A.,
 RA Sharp J., Wheeler R., Kusumi K., Moie S., Liu W., Soares M.B.,
 RA Bonaldo M.F., Hirvasniemi A., de la Chapelle A., Gilliam T.C.,
 RA Lehesjoki A.-E.;
 RT "The neuronal ceroid lipofuscinoses in human EPMR and mnd mutant mice
 RT are associated with mutations in CLN8.";
 RL Nat. Genet. 23:233-236(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary gland;
 RX MEDLINE=22388257; Pubmed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
 CC reticulum and ER-Golgi intermediate compartment (EXCIC) (By
 CC similarity).
 CC -|- DISEASE: Defects in CLN8 are the cause of the phenotype motor
 CC neuron degeneration (mnd). Mnd is a naturally occurring mutant it
 CC is characterized by progressive motor system degeneration. It has
 CC intracellular autofluorescent inclusions similar to those seen in
 CC human CLN8.
 CC -|- SIMILARITY: Contains 1 TLC (TRAM/LAG1/CLN8) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF125308; AAF04462.1; -.
 DR EMBL; AF125307; AAF04461.1; -.
 DR EMBL; BC021625; AAH21625.1; -.
 DR MGD; MGI:1349447; Cln8.
 DR InterPro; IPR006634; TLC.
 DR SMART; SM00724; TLC; 1.
 DR PROSITE; PS50922; TLC; 1.
 KW Transmembrane; Endoplasmic reticulum; Neuronal ceroid lipofuscinosis.
 FT TRANSMEM 26 46 POTENTIAL.
 FT TRANSMEM 71 91 POTENTIAL.
 FT TRANSMEM 103 123 POTENTIAL.
 FT TRANSMEM 131 151 POTENTIAL.

```
FT TRANSMEM 225 245 POTENTIAL.
FT DOMAIN 62 262 TLC.
FT SITE 285 288 ER-RETRIEVAL SIGNAL.
SQ SEQUENCE 288 AA; 33108 MW; A11225DE3CD7DEA5 CRC64;

Query Match
Best Local Similarity 25.9%; Score 100.5; DB 1; Length 288;
Matches 45; Conservative 17; Mismatches 75; Indels 37; Gaps 7;

QY 47 GHSRWFVLVRVLSLFIGAIVAVHSAEWEVGTNTNTSYKAFSAARYTARVGLVGL 106
DQ 98 GQCNWCWFHITTAGFFFEFENVAVHLS-----NLFERTFDLFLVHHLFAFLGFL 147
QY 107 GINITLTGTPVHQNET---IDYNEQFT---WRLKENYAAEYANALEKGLPDPVLYAEK 160
DQ 148 GSAINLRAG-HYLAHTLTLENTSTPFTCISWML-----LKAGWSDSLFWKANO 194
QY 161 FTSPSPCGLYHQYHLAGHYASATLWVAFCFW--LLSNVLSTPAPLYGGLALT 212
DQ 195 WL-----MIHMFGRMILTYHMMWVCFCHMDALTSTSLHPSHWALFLFGALLT 242

RESULT 10
CYB_CARPL STANDARD; PRT; 381 AA.
AC P34866;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome b.
GN MTCYB OR COB OR CYTB.
OS Carcharhinus plumbeus (Sandbar shark).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Galeoidea; Carcharhiniformes;
OC Carcharhinidae; Carcharinus.
OX NCBI_TaxID=7808;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92252907; PubMed=1579163;
RT "Rates of mitochondrial DNA evolution in sharks are slow compared with mammals."
RL Nature 357:153-155 (1992).
CC -!- FUNCTION: Component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex), which is a respiratory chain that generates an electrochemical potential coupled to ATP synthesis (By similarity).
CC -!- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or b562) is low-potential and absorbs at about 562, and heme 2 (or BH or b566) is high-potential and absorbs at about 566 (By similarity).
CC -!- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b, cytochrome c1 and the Rieske protein (By similarity).
CC -!- SIMILARITY: Belongs to the cytochrome b family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
DR EMBL; L08032; AAA31703.1; -
DR InterPro; IPR005798; Cytb_b6_c.
DR InterPro; IPR005797; Cytb_b6_N.
DR Pfam; PF00032; cytochrome_b_c1; 1.
DR Pfam; PF00033; cytochrome_b_N; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE; PS00193; CYTOCHROME_B_OO; 1.
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW Heme.
```

```
FT METAL 84 84 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 183 183 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 197 197 IRON 2 (HEME B566 AXIAL LIGAND).
SQ SEQUENCE 381 AA; 43366 MW; 8A429923A8F17A11 CRC64;

Query Match
Best Local Similarity 23.4%; Score 100.5; DB 1; Length 381;
Matches 64; Conservative 46; Mismatches 95; Indels 69; Gaps 15;

QY 27 VILVFLALAAFL-LILPGIRGSRN---FWLVRVLSL-----FIGAIVAVHSAEW 76
DQ 118 VILFLMLATAFVGYP-----WGQMSFWGATVITNLSAPPIYIGDMTV-----QW 164
QY 77 FVGTVN---TNTSYKAFSAARYTARVGL-LVGLGINITLTGTP--VHQLNETIDYNEQ 129
DQ 165 IWGFSVDNATLTRPFAFHLPLFLIALTIHLLFLETGSNNPLGINSDDAKISFHPY 224
QY 130 FTWR-----LKENYAAEYANALEKGLPDPVLYAEKFTSPSPCGLYHQYHLAGHYASAT 183
DQ 225 PSYKDLGFFVMIFFLAVFALFMPNLGD-----AENFIPANP--LVTPPHIREW----- 273
QY 184 LWVAFCEFWLLSNVLSTPAPLYGGLALTGTGAFALFGVFALASISSVPLCPRLGSSALT 243
DQ 274 -----YFLFAVAILRSLPNKLGVTAL-----LFSIFILM-----LVPLHTSKORS 315
QY 244 TQYGA-----FWTLATGVLCFLGCAVVSLOVY 273
DQ 316 TIFRPMQIFFWLLVANSIILTWIGQVPEQPF 349

RESULT 11
CYB_SCYCA STANDARD; PRT; 381 AA.
AC 079413;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome b.
GN MTCYB OR COB OR CYTB.
OS Scyliorhinus canicula (Spotted dogfish) (Spotted catshark).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Galeoidea; Carcharhiniformes;
OC Scyliorhinidae; Scyliorhinus.
OX NCBI_TaxID=7830;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Muscle;
RX MEDLINE=98393590; PubMed=9725850;
RT Delarbre C., Spryut N., Delmarre C., Gallut C., Barriol V., Janvier P., laudet V., Gachelin G.;
RT "The complete nucleotide sequence of the mitochondrial DNA of the dogfish, Scyliorhinus canicula."
RL Genetics 150:331-344 (1998).
CC -!- FUNCTION: Component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex), which is a respiratory chain that generates an electrochemical potential coupled to ATP synthesis (By similarity).
CC -!- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or b562) is low-potential and absorbs at about 562, and heme 2 (or BH or b566) is high-potential and absorbs at about 566 (By similarity).
CC -!- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b, cytochrome c1 and the Rieske protein (By similarity).
CC -!- SIMILARITY: Belongs to the cytochrome b family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```

CC      or send an email to license@isb-sib.ch).
CC
CC      -----
DR      EMBL; Y16067; CAA76031.1; -.
DR      PIR; T11312; T11312.
DR      InterPro; IPR005798; Cytb_b6_C.
DR      InterPro; IPR005797; Cytb_b6_N.
DR      Pfam; PF00032; cytochrome_b_C; 1.
DR      Pfam; PF00033; cytochrome_b_N; 1.
DR      PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR      PROSITE; PS00193; CYTOCHROME_B_CO; 1.
KM      Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KM      Heme.
FT      METAL      84      84      IRON 1 (HEME B562 AXIAL LIGAND) .
FT      METAL      98      98      IRON 2 (HEME B566 AXIAL LIGAND) .
FT      METAL      183     183     IRON 1 (HEME B562 AXIAL LIGAND) .
FT      METAL      197     197     IRON 2 (HEME B566 AXIAL LIGAND) .
SQ      SEQUENCE   381 AA; 43103 MW; C4B457072A56D0C5 CRC64;

```

[illegible]

```

RESULT 12
YGTU_ECOLI
ID_YGTU_ECOLI STANDARD; PRT; 414 AA.
AC P42602;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical symporter YgtU.
GN YGTU OR B3089 OR Z4442 OR ECS3971.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirtpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).

```

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935: PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimailanta E.T., Potamuscis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blatner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RT Nature 409:529-533(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurowaka K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakiwa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RNA Res. 8:11-22(2001).
RN [4]
RP SEQUENCE OF 268-447 FROM N.A.
RC STRAIN=K12 / W3110;
RA Mizobuchi K.;
RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE SODIUM-DICARBOXYLATE SYMPORTER FAMILY
CC (SDF, TC 2.A.23). STRONG, TO H.INFLUENZAE H1545.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
CC or send an email to license@ebi.ac.uk).

CC	EMBL	U18997; AAA57891.1; -	-
DR	EMBL; AE000391; AAC76124.1; -	-	-
DR	EMBL; AE005538; AAG58222.1; -	-	-
DR	EMBL; AP002564; BAB37394.1; -	-	-
DR	EMBL; D13328; -; NOT_ANNOTATED_CDS.	-	-
DR	PIR; B85970; B85970.	-	-
DR	PIR; C91125; C91125.	-	-
DR	PIR; F65097; F65097.	-	-
DR	EcGene; EG12732; y9jU.	-	-
DR	InterPro; IPR001991; Na/d1CO_symport.	-	-
DR	Pfam; PF00375; SDF; 1.	-	-
DR	PRINTS; PR00173; EDTRANSPORT.	-	-
DR	PROSITE; PS00713; NA_DICARBOXYL_SYMPT_1; FALSE_NEG.	-	-
DR	PROSITE; PS00714; NA_DICARBOXYL_SYMPT_2; FALSE_NEG.	-	-
KW	Hypothetical protein; Transmembrane; Transport; Complete proteome.	-	-
FT	TRANSMEM	16	36
FT	TRANSMEM	46	66
FT	TRANSMEM	84	104
FT	TRANSMEM	143	163
FT	TRANSMEM	180	200
FT	TRANSMEM	218	238
FT	TRANSMEM	300	320
FT	TRANSMEM	332	352
FT	TRANSMEM	359	379
SEQUENCE	414 AA; 43477 MW; 47713C1DB77B1E38 CRC64;	-	-

Query Match	6.0%;	Score 100.5;	DB 1;	Length 414;
Best Local Similarity	22.3%;	Pred. No. 0.77;		
Matches 74;	Conservative 55;	Mismatches 134;	Indels 69;	Gaps 16;
QY	17	AAGFSVPLLIYLIVFLALAA	FLLLPGIRGHSRWFML-V	RVLLSLFT----GAELVAVH 71
Db	42	AVGLGLTLFVGALKAVAPIL	VLMLVMASIANHQGKTNIR	PIFLFYLLGLTFSALLAAV 101
QY	72	PSAEWFGVTNTNTSYKAFS	ARVTARVGLLVGLEGINIT	LGTGPVHQINETIDYNEQFT 131
Db	102	PSFAF-----PSTLHLSS	SAGDISPPSGIIEVMRGLV	MSVSNPIDALKG-NYIGILV 154
QY	132	WRLKENYAAYANALEKGL	----PDVLYLAEKFTPSS	PCGILHQYHLAGHYAS----- 181

```

Db      155 WAIGLGFALRHGNETKLVNDMSNAVTFMWKLVIRFAPIGIF-----GLVSTLATTTG 208
QY      182 -ATLM-----VAFCFWLLS---NVLL-----STPAPLYGGLALLTTGAFALFGVFA 223
Db      209 FSTLMGYAQLLVAVGCMMLLVAVNPLVWVKIRNPFPLV-LLCLREGVYAFTRSS 267
QY      224 LASIS-SVPLC-----PLRLGSSALTYGGAFFWTLATGVLCFLGAVVSLQY 272
Db      268 AANIPVMALCEKXNLDRDYSVSIPLGATINMAGAITITVTLAAVNTLG-----IPV 322
QY      273 VRPSALRTLLDQSAKDC---SQERGGSPLL 300
Db      323 DLPTAL--LLSVVASLACGASGVAGGSLILLI 352

RESULT 13
CYB_RANNI
ID      CYB_RANNI      STANDARD;      PRT;      380 AA.
AC      Q9T6R9;
DT      10-OCT-2003 (Rel. 42, Created)
DT      10-OCT-2003 (Rel. 42, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Cytochrome b.
GN      MTCYB OR COB OR CYTB.
OS      Rana nigromaculata (Japanese pond frog).
OG      Mitochondrion.
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Ranidae; Rana.
RN      NCB1_TaxID=8409;
RP      [1]
RA      SEQUENCE FROM N.A.
RT      Lee J.-E., Lee H.-Y., Yang S.-Y.;
RT      "Genetic relationships of mitochondrial cytochrome b gene among six
RT      Korean Rana species."
RL      Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: Component of the ubiquinol-cytochrome c reductase
CC      complex (complex III or cytochrome b-c1 complex), which is a
CC      respiratory chain that generates an electrochemical potential
CC      coupled to ATP synthesis (By similarity).
CC      -1- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or
CC      b562) is low-potential and absorbs at about 562, and heme 2 (or BH
CC      or b566) is high-potential and absorbs at about 566 (By
CC      similarity).
CC      -1- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
CC      cytochrome c1 and the Rieske protein (By similarity).
CC      -1- SIMILARITY: Belongs to the cytochrome b family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AF205087; AAF17086.1; -.
DR      InterPro; IPR005798; Cytb_b6_C.
DR      InterPro; IPR005797; Cytb_b6_N.
DR      Pfam; PF00032; cytochrome_b_C; 1.
DR      Pfam; PF00033; cytochrome_b_N; 1.
DR      PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR      PROSITE; PS00193; CYTOCHROME_B_CO; 1.
KW      Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW      Heme.
FT      METAL      84      84      IRON 1 (HEME B562 AXIAL LIGAND).
FT      METAL      98      98      IRON 2 (HEME B566 AXIAL LIGAND).
FT      METAL      183     183     IRON 1 (HEME B562 AXIAL LIGAND).
FT      METAL      197     197     IRON 2 (HEME B566 AXIAL LIGAND).
SQ      SEQUENCE      380 AA; 42204 MW; 1453BD8F0754F84A CRC64;

Query Match      6.0%; Score 99.5; DB 1; Length 380;
Best Local Similarity      22.5%; Pred. No. 0.84;

```

```

Matches      61; Conservative      45; Mismatches      102; Indels      63; Gaps      14;
QY      27 VILVFLAASFL-LIIPGIRGHSRW----FWLVFVLSL-----FIGAIVAVHFSAEW 76
Db      118 VILLFLVATAFVGXYLP-----WGQMSFWGATVITNLSAPYIGPDV-----QW 164
QY      77 FVG--TVNNTSYKAFSAARVAV---GLLVLEGINITLTGTPVHQLNETD----- 125
Db      165 IWGFSVNSTLTFRFFHFILPFIIAASMIHLFLHQTGSSNPT-GLNSNIDKVSFHP 223
QY      126 ---YNEQFTWLKENTAAEYANALEKGLPDPVLYLAEKFTPSSPCGLYHQYHLAGHYASA 182
Db      224 YFSYKDLGFVIMLGALASLSTFAPNLLGP-----DNFTPANP-LVTPPHIKPEW--- 273
QY      183 TLMVAFCFWLLSNVLSLSTPAPLYGGLALLTTGAPALFGVPALASISSVPLCPRLGSSAL 242
Db      274 -----YFLPAYAILKRSIPNKLGGVLAAL---LSIMVLFMLPIHTSKRSIMFRPIAK 323
QY      243 TTQYGAFFWTLATGVLCFLGAVVSLQYV 273
Db      324 T-----FFWALIANTAALLTWIGGQPVEDPFI 349

```

```

RESULT 14
NU5C_ORYSA
ID      NU5C_ORYSA      STANDARD;      PRT;      734 AA.
AC      P12129;
DT      01-OCT-1989 (Rel. 12, Created)
DT      01-OCT-1989 (Rel. 12, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      NAD(P)H-quinone oxidoreductase chain 5, chloroplast (EC 1.6.5.-)
DE      (NAD(P)H dehydrogenase, chain 5) (NADH-plastoquinone oxidoreductase
DE      chain 5).
GN      NDHF.
OS      Oryza sativa (Rice).
OG      Chloroplast.
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC      Ehrhartoideae; Oryzaceae; Oryza.
OX      NCB1_TaxID=4530;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Nipponbare;
RX      MEDLINE=89364698; PubMed=2770692;
RA      Hiratsuka J., Shimada H., Whittier R., Ishibashi T., Sakamoto M.,
RA      Mori M., Kondo C., Honji Y., Sun C.-R., Meng B.-Y., Li Y.-Q.,
RA      Kanno A., Nishizawa Y., Hirai A., Shinozaki K., Sugita M.;
RT      "The complete sequence of the rice (Oryza sativa) chloroplast genome:
RT      intermolecular recombination between distinct trna genes accounts for
RT      a major plastid DNA inversion during the evolution of the cereals.";
RL      Mol. Genet. 217:185-194(1989).
CC      -1- CATALYTIC ACTIVITY: NAD(P)H + plastoquinone = NAD(P)H(+) +
CC      plastoquinol.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X15901; CAA33950.1; -.
DR      PIR; JQ0286; DERZNS.
DR      Gramene; P12129; -.
DR      InterPro; IPR003916; NADhub_oxred5.
DR      InterPro; IPR001750; Oxidored_q1.
DR      InterPro; IPR002128; Oxidored_q1_C.
DR      InterPro; IPR001516; Oxidored_q1_N.
DR      Pfam; PF00361; oxidored_q1; 1.
DR      Pfam; PF01010; oxidored_q1_C; 1.
DR      Pfam; PF00662; oxidored_q1_N; 1.
DR      PRINTS; PR01434; NADHDHGNASES.

```


KW Oxidoreductase; NAD; NADP; Quinone; Plastoquinone; Chloroplast.
SQ SEQUENCE 734 AA; 82597 MW; BFEFEE20A01632B2 CRC64;

Query Match 5.9%; Score 99; DB 1; Length 734;
Best Local Similarity 22.9%; Pred. No. 1.9;
Matches 72; Conservative 46; Mismatches 88; Indels 108; Gaps 20;

```
QY 23 PLIVIVLFLAASFLI-----LPGRGHSRWFVLVRV-----LLSLFIGAIVAVHS 73
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 91 PLTSMILITTVGLIVLYSDYMSHDEGLRFEVYISFNTSMGLVTSNLIQIVFF 150
    ::::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 74 AE-----WEVGTNTNTSYKAFSAARYTARV---GLLVGLEGINITLTGTPVH 118
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 151 WELVGMCSYLLIGFWETRPAAACQKAF---VTNRVGDGGLLGLGF-FWITG--- 201
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 119 QLNETIDYNEQFTWRLEKENYAE-----YANALEKG-----LPPDVL 155
    ::::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 202 ---SLEFRDLF--KIANNWIPNNEINSLTLICAFLLFLGAVAKSAQFPLHVLPDAM- 254
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 156 YLAKEFTPSSPCGLYHQYHLAGHYASATLWVAFCEWLSNVLLSTPAPLYGGLALLTT-- 213
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 255 ---EGPTPIIS--ALIR-----ATTWVAGIFLIARL-----PLFISLPLIMSFT 294
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 214 ---GAPALF--GVFALA-----SISVPLCPRLGSSALLTQYGAFFW--VTIA 255
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 295 SLIGTLTLFLGATLALAQDRIKSLAYSTMSQLGYMLALGIGS---YQALPHLITHA 350
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 256 TGVLCFLG-GAVV 268
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 351 YSKALLFLGSGSVI 364
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||
```

```
RESULT 15
CYB_DROME STANDARD; PRT; 378 AA.
ID CYB_DROME STANDARD; PRT; 378 AA.
AC P18935;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome b.
GN MT:CYT-B OR COB OR CYTB.
OS Drosophila melanogaster (Fruit fly).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bretagne;
RX MEDLINE=88212147; PubMed=3130291;
RA Garesse R.;
RT "Drosophila melanogaster mitochondrial DNA: gene organization and
RT evolutionary considerations.";
RL Genetics 118:649-663(1988).
CC -|- FUNCTION: Component of the ubiquinol-cytochrome c reductase
CC complex (complex III or cytochrome b-c1 complex), which is a
CC respiratory chain that generates an electrochemical potential
CC coupled to ATP synthesis (By similarity).
CC -|- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or
CC b562) is low-potential and absorbs at about 562, and heme 2 (or BH
CC or b566) is high-potential and absorbs at about 566 (By
CC similarity).
CC -|- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
CC cytochrome c1 and the Rieske protein (By similarity).
CC -|- SIMILARITY: Belongs to the cytochrome b family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

CC or send an email to license@isb-sib.ch.

```
CC -----
CC EMBL; M37275; AAA69714.1; -.
DR EMBL; U37541; AAC47822.1; -.
DR PIR; S01190; S01190.
DR FlyBase; FBgn0013678; mt:Cyt-b.
DR InterPro; IPR005798; Cytb_b6_C.
DR InterPro; IPR005797; Cytb_b6_N.
DR Pfam; PF00032; cytochrome_b_c1_1.
DR Pfam; PF00033; cytochrome_b_N; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE; PS00193; CYTOCHROME_B_OO; 1.
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW Heme.
FT METAL 84 84 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 183 183 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 197 197 IRON 2 (HEME B566 AXIAL LIGAND).
SQ SEQUENCE 378 AA; 43224 MW; 6CB9067643ABCE83 CRC64;
```

Query Match 5.9%; Score 98.5; DB 1; Length 378;
Best Local Similarity 20.8%; Pred. No. 1;
Matches 58; Conservative 52; Mismatches 114; Indels 55; Gaps 13;

```
QY 24 LLIVIVLFLAASFL-LILPGIRGHSRW---FWLVRVLLSL-----FIGAIVAVHS 73
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 115 LIGVITLFLVMTAFMGVLP-----WGQMSFWVATVITNLLYALPYLGMDLV----- 162
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 74 AEWFGV--TVNTNTSYKAFSAARYTARVGL---LVGLEGINITLTGTPVHQNETID--- 125
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 163 -QWLMGFAVDNATLTRFTTFHFLPFIVLAMTMHLLFHQTSNNPI-GLNSNDIKIP 220
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 126 YNEQFTWRLEKENYAAEVANALEKGLPDP-VLYLAKEFTPSSPCGLYHQYHLAGHYASATL 184
    ::::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 221 FHPYFTFKDIYGFIVMIFILISLVLSNLLGDPDNFIPTPL-----VTPAHI 269
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 185 WVAFCFWLLSNVLLSTPAPLYGGLAL-LTGAFAFLGVFALASISSVPLCPRLGSSALT 243
    ::::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 270 QPEWYFLFAVALIRISPNKLGIVIALVLAIMLIPFYNLSKRGIOFYYPIN----- 322
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 244 TQYGAFFWTATGVLCLFLGAVVSLQYVRPSALRTLL 282
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 323 ---QVMFWSMLVTVILTWIGARPVEEPYVLIGQILTVV 358
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||
```

Search completed: February 23, 2004, 19:17:59
Job time : 51 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 23, 2004, 19:10:51 ; Search time 80 Seconds
(without alignments)
1262.073 Million cell updates/sec

Title: US-09-936-456-2
Perfect score: 1666
Sequence: 1 MTLWNGVLPFPYQPRHAGF.....GDP LHKQALPDLKICITTNL 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- SPTREMBL_25:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_virus:*
 - 16: sp_bacteriophage:*
 - 17: sp_archaeophages:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1312	78.8	320	11 Q9D311	Q9D311 mus musculus
2	987	59.2	320	11 Q9D7U8	Q9D7U8 mus musculus
3	883	53.0	483	4 Q8N6K9	Q8N6K9 homo sapien
4	851.5	51.1	341	11 Q8VE49	Q8VE49 mus musculus
5	697.5	41.9	298	4 Q96M14	Q96M14 homo sapien
6	375	22.5	474	5 Q9VUQ7	Q9VUQ7 drosophila
7	254.5	15.3	363	5 Q9V441	Q9V441 drosophila
8	127.5	7.7	391	16 Q34929	Q34929 bacillus su
9	121.5	7.3	461	16 Q9KEG8	Q9KEG8 bacillus ha
10	115	6.9	467	16 Q8Z128	Q8Z128 salmonella
11	115	6.9	815	16 Q9ZBJ1	Q9ZBJ1 streptomyc
12	114.5	6.9	446	17 Q974G1	Q974G1 sulfolobus
13	113.5	6.8	792	16 Q8Z1V7	Q8Z1V7 streptomyc
14	113	6.8	467	16 Q8ZK36	Q8ZK36 salmonella
15	113	6.8	509	8 Q47547	Q47547 chlamydomon
16	109	6.5	495	17 Q8TCN5	Q8TCN5 methanosarc

17	108.5	6.5	512	2 Q936C7	Q936C7 streptomyc
18	108	6.5	469	17 Q8TIE5	Q8TIE5 methanosarc
19	107.5	6.5	289	16 Q8FUP6	Q8FUP6 brucella su
20	107.5	6.5	378	8 Q9TEL8	Q9TEL8 ceratitlis c
21	107.5	6.5	544	16 Q9RJ46	Q9RJ46 streptomyc
22	107	6.4	380	8 Q7YHJ2	Q7YHJ2 aliofooca ca
23	107	6.4	395	16 Q88UH3	Q88UH3 lactobacill
24	106.5	6.4	445	17 Q96X94	Q96X94 sulfolobus
25	106.5	6.4	731	8 Q85FH9	Q85FH9 adiantum ca
26	106	6.4	380	8 Q7YE82	Q7YE82 chromis cya
27	106	6.4	495	17 Q9BP94	Q9BP94 methanosarc
28	106	6.4	573	16 Q89QM0	Q89QM0 bradyrhizob
29	105.5	6.3	326	16 Q8Y2U9	Q8Y2U9 ralsionia s
30	105.5	6.3	372	8 Q21685	Q21685 pseudocarch
31	105.5	6.3	647	16 Q83AT2	Q83AT2 coxiella bu
32	104.5	6.3	349	8 Q03547	Q03547 rhytidopone
33	104.5	6.3	380	8 Q7YE23	Q7YE23 plectroglip
34	104.5	6.3	380	8 Q7YB05	Q7YB05 pomacentrus
35	104	6.2	364	2 Q83U30	Q83U30 streptomyc
36	104	6.2	703	8 Q32532	Q32532 lithachne p
37	103.5	6.2	349	8 Q03514	Q03514 rhytidopone
38	103.5	6.2	375	8 Q9T9G3	Q9T9G3 pupa strigo
39	103.5	6.2	502	16 Q82NC5	Q82NC5 streptomyc
40	103.5	6.2	740	8 Q8W6H5	Q8W6H5 psilotum nu
41	103	6.2	380	8 Q7YE87	Q7YE87 chromis atr
42	103	6.2	380	8 Q7YE75	Q7YE75 chromis vir
43	103	6.2	403	2 Q9X4K3	Q9X4K3 pseudomonas
44	102.5	6.2	288	11 Q8BNW2	Q8BNW2 mus musculu
45	102.5	6.2	322	16 Q8YV97	Q8YV97 anabaena sp

ALIGNMENTS

Q9D311	PRELIMINARY;	PRT;	320 AA.
ID	Q9D311		
AC	Q9D311		
DT	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DT	01-OCT-2002 (TREMBLrel. 22, Last annotation update)		
DE	9030623N16R1K protein (RIKEN CDNA 9030623N16 gene).		
GN	9030623N16R1K.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Colon;		
RX	MEDLINE=21085660; PubMed=11217851;		
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,		
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,		
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,		
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,		
RA	Kadota K., Matsuda H.A., Ashburner M., Batcalov S., Casavant T.,		
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,		
RA	Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,		
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,		
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,		
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,		
RA	Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,		
RA	Gustincich S., Hill D., Hofmann M., Mazzarelli J., Mombaerts P.,		
RA	Lyons P., Marchionni L., Mashima J., Mazzaletti J., Mombaerts P.,		
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,		
RA	Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,		
RA	Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,		
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,		
RA	Hayashizaki Y.,		
RT	"Functional annotation of a full-length mouse CDNA collection."		
RL	Nature 409:685-690(2001).		
RN	[2]		
RP	SEQUENCE FROM N.A.		


```
Qy      247 GAAFWVTLATGVLCLFLGGAVWSLQYVRPSARLTLLDSAK-----DCSQERG 295
        |||::||::||::||::||::||::||::||::||::||::||::||::||
Db      248 GPAFWTLITGLLCVLLGLAMAVAHKMQPHRIKAFFNQSDVEDPMLEWSPEEGG 301
```

RESULT 4

ID	Q8VE49	PRELIMINARY;	PRT;	341 AA.
AC	Q8VE49;			
DT	01-MAR-2002 (TrEMBLrel. 20, Created)			
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	Hypothetical protein.			
GN	BC019755.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxId=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Strausberg R.;			
RL	Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC019755; AAH19755.1; -			
DR	MGI; MGI:2384861; BC019755.			
KW	Hypothetical protein.			
SQ	SEQUENCE 341 AA; 37595 MW; CC16D7AD103A7276 CRC64;			

Query Match	51.1%;	Score 851.5;	DB 11;	Length 341;
Best Local Similarity	56.7%;	Pred. No. 4,4e-62;		
Matches 166;	Conservative 40;	Mismatches 82;	Indels 5;	Gaps 2;

[illegible]

RESULT 5

ID	Q96M14	PRELIMINARY	PRT	298 AA.
AC	Q96M14;			
DT	01-DEC-2001	(TREMBLrel. 19, Created)		
DT	01-DEC-2001	(TREMBLrel. 19, Last sequence update)		
DT	01-OCT-2002	(TREMBLrel. 22, Last annotation update)		
DE	Hypothetical protein FLJ32334.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Carnivora; Homini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
R2	SEQUENCE FROM N.A.			
RC	TISSUE=Prostate;			
RA	Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,			
RA	Hotuta T., Hirooka S., Murakawa K., Takiguchi S., Kusano J.,			
RA	Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,			
RA	Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,			
RA	Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,			

RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Magatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,
RT "NEDO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.

RP SEQUENCE FROM N.A.

RA Strausberg R.; submitted (JAN-2002) to the EMBL/GenBank/DBJ databases
 RL EMBL; AK056896; BAB71304.1; -
 DR EMBL; BC020841; AAH20841.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 298 AA; 33110 MW; D9B346C25D142E1D CRC64;

SQ SEQUENCE 298 AA; 33110 MW; D9B346C25D142E1D CRC64;

Query Match	41.9%	Score 697.5;	DB 4;	Length 298;
Best Local Similarity	48.6%;	Pred. No. 1.9e-49;		
Matches 143;	Conservative 32;	Mismatches 66;	Indels 53;	Gaps 4;

```

QY      9 PFYQPQRHAGFSVPLILVILVFLALAAPELLPGIRSHSRMFWLVRVLSLFIGAEIV 68
      ||| | : : : || : : : || : : : || : : : || : : : || : : : || : : :
Db      9 PFYAGPKPTFPMDDTLASIMIFLTALATFIVILPGIRGTRRLFWLVRVTSLSFIGAAL 68

```

QY	69	AVHFSAEWVGTVNTNTSYKAFSAARVTARVGLLVEGINITLTGTPVHQNETIDYNE	128
		:	
Db	69	-----GTPVQQLNETINYN	83

```

QY      129 QFTWRLKENYA AEYANALAEKGLPDVLYLAEKFTSPCGLYHQYHLAGHYASATLWVAF 188
      :||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      84 EFTWRLGGEYAAEYAKALAEKGLPDVLYLAEKFTSPCGLYHQYRLAGHYTSAMLWVAF 143

```

QY 189 CFMLLSNVLLSTPAPLYGGLALLTGAFALFGV--FALASISSVPLCPRLRGSSALTTOY 246
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
Db 144 LCMILLANVMISMPVLYGGYMLLATGIQLLALLEFSMATSLTSP-CPHLHGASVLHTHH 202

```
QY      247 GAAFWTLATGVLCLFLGGAVSLQYVRPALRTLLDOSAK-----DCSQERG 295
          |||:|||||::|||::|||::|||::|||::|||
Db      203 GPAFWTLLTTGLLCVLLGLAMVAHRMQPHRLKAFENQSVDEDPMLESPEEGG 256
```

RESULT 6

ID	Q9VUQ7	PRELIMINARY;	PRT;	474 AA.
AC	Q9VUQ7; Q95RU0;			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	CG4482 protein (JD27791p).			
GN	BG:DS01219.1 OR CG4482 OR CG15268.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			

[1]

RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10721132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoekstra R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]

RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Change M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,
 RT "Sequencing of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A.
 RA Mistra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Ciamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,
 RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]

RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]

RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]

RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE003645; AAF53425.2;
 DR EMBL: AY061346; AAL28894.1;
 DR FlyBase; FBgn0028528; BG:DS01219.1.
 SO SEQUENCE 474 AA; 53348 MW; C48A5008304CDE64 CRC64;

Query Match 22.5%; Score 375; DB 5; Length 474;
 Best Local Similarity 34.4%; Pred. No. 1.3e-22;
 Matches 93; Conservative 46; Mismatches 115; Indels 16; Gaps 9;
 QY 22 VPLIIVLWFLAASFLILPGIRGSRFWLVRLVLSLFGAEIYAVHSAEYVGTV 81
 DB 29 VSIIVSVLFAFTRVYAVFLVPPGVR-KQKFTTSTVLSLFGVLITRLGSAHVYAH 87
 QY 82 NNTSYKAFSAARVATRVGLVGLGINTLTGTPV-HQNETIDYNEOFTWRLKENYAA 140
 DB 88 TTIADPYKAFSREKLPARIGTHIGLMHVNVTITAPIGMTTPPDIDYNERFTWEGANDMSA 147
 QY 141 EYANALEKGLPDPVLYLAKEFTPSSPCGLY-HQYHLAGHYASATLWAFCEWLSNVLLS 199
 DB 148 NTRHALQRLPFPILITVAEYFSLGREGFSWGGYRAAGYFASIMTWASLAWLWMLL- 206
 QY 200 TPAPLYGLALLTTGAFAL---FGVFALASISSVPLCL--RLGSSALTQYGAFAFWTL 254
 DB 207 IAVPRYGAYMKALTGALLVCTVGYHCL-----LPKRPLSIHLEGRLEFHFQWCYWLVL 261
 QY 255 ATGVLCFLGGA VVS-LQYVPSALRTILD 283
 DB 262 VAGTLC-FIAGVLISIIDLVWPHFTSTYLE 290

RESULT 7
 QYV441 PRELIMINARY; PRT; 363 AA.
 ID Q9V441;
 AC Q9V441;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE BG:DS01219.1 protein.
 GN BG:DS01219.1 OR CG4482 OR CG15268.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Maassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=99403001; PubMed=10471707;
RA Ashburner M., Miera S., Roote J., Lewis S.E., Blazek R., Davis T.,
RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whiteaw K.,
RA Celniker S., Rubin G.M.;
RT "An exploration of the sequence of a 2.9-Mb region of the genome of
RT *Drosophila melanogaster*: the *Adh* region.";
RL Genetics 153:179-219(1999).
DR EMBL; AE003645; AAG22432.1; -;
DR EMBL; AE003411; AAF44903.1; -;
DR FlyBase; FBgn0028528; BG:DS01219.1.
SQ SEQUENCE 363 AA; 41170 MM; 1912DA313B6F8446 CRC64;

Query Match	15.3%;	Score 254.5;	DB 5;	Length 363;
Best Local Similarity	35.3%;	Pred. No. 8e-13;		
Matches 65; Conservative	30;	Mismatches 74;	Indels 15;	Gaps 8

```
QY      108 INITLTGPEV-HQLNETIDYNEQFWRLKENYAAYANALEKGIJDOPVLYLAEKFTPPSGP 166
       :|:|:| | : : ||||| : : :|:|:|:| | :|:|:| :
Db      3 UNVTTLTALPICMWTTPPIDIDYNERFTEMGANDMSANYRHALORGI PPILITLVAEYFSLGE 62
```

QY 167 CGLY-HQYHLAGHYASATLMVAFCEFMILSNVLSTPAPLYGLALTTGAPAL---FGVF 222
: || ||::|| ||: ||||::|| || ||: ||
Db 63 GFSMGQVPAAQYFASIMCMASLASHNIMLNL-L-IAVPRYGATMKALTGALLVCTTGVGH 121

QY 223 ALASISSVPLCP-RLGSSALTTQYGAAFWMTLATGVLCLEFGAGAVS-IQYVRPSALR 279
| : | | : : | : | : | : | :
Db 122 CL-----LPKRPISIHLEGGRLEFHFGMCWMLVLVAGILC-FIAGVLSISIDLVMPHTFS 175

QY	280	TLLD	283
Db	176	TVLE	179

RESULT 8
034929
ID 034929 PRELIMINARY; PRT: 391 AA.

DT	01-JAN-1998 (TREMblrel. 05, Created)
DT	01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT	01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE	YFKF protein.

OS *Bacillus subtilis*.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Bacillus*.
OX NCBI TaxID=1423;

RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;

RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchet S.,
RA Bottiss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell E., Capuano V., Carter N.M.,

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmeson P.T.,
Garcia J.R., O'Connell L.T., O'Connell N.T., O'Connell N.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA

RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Hatech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,

RA Joris E., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Mediglic C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivaolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Sedor S.U., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,
RA Takeuchi M., Takamashi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*

RT Subtilisin.²
 RL Nature 390:249-256 (1997).
 RN {2}
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AC327;
 RA Sekiguchi J., Yamamoto H., Uchiyama S., Fajar A.;
 RI Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.

DR EMBL; Z39108; CAB12620.1; -.
DR EMBL; D83967; EAA23396.1; -.
DR PIR; B69808; B69808.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub_transporter.

DR PROSITE; PS50850; MFS; 1.
KW Transmembrane; Complete proteome.
SQ SEQUENCE 391 AA; 42690 MW; F909AD127C010ADD CRC64;

Query Match	7.7%	Score 127.5;	DB 16;	Length 391;
Best Local Similarity	25.3%;	Pred. No. 0.026;		
Matches 75; Conservative	33;	Mismatches 105;	Indels 83;	Gaps 16;

```
QY      3 LMNGVL---EFYPQPRHAAGFSVPLLIIV--LIVFLLAALSLIIIDGIRGHSRWELVRV 57
        | : ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     48 LYIGVLLASPMEAPLRKLGFK-PLIIVGGSIIVLSLGFIML-----QSVWVF---- 96
```

Qy 58 LLSLFIGAIVVHFSAEVFVGTNTNITSYKAFSARVTARVGELVLEGINITLTGTPV 11
||| : ||| : | | : |
Db 97 LIRLFIGIGDHMLHFSTGIW----TSMSSKQ-NRGKNLSIYGSLFGL-GFAAGPFMVPL 150

```

QY      118 HOLNETIDY-----NEQFTWRLKENTAAEYANALEKGLPDPVLYIAEKFTPSSP----- 166
      :|:::|-----|:::|
Db      151 VKLSPSLPFIIVSGCISLJFAMLF-----VFLLQNAVYPETSPHETKSD 197

```

```

QY      167  --CGLYHQHLAGHYASATIMWAFCEFWLSNVLSTPAPLYGGALLTTGAFALGVEFAL  22
          ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      192  NSFRRFYQAMLFG-----WVAF-----WETFGYGFLETALNGSFPTVALRLG  233

```

```
QY      225 ASISSVP--LCPRLRGS-----SALTTOYGA--AFWVTLATGVLCIFLGAVNS 263
          :  : |         : ||       :  : |   |   |   |   |   |   |
Db      234 ISVDVAAILPAPALGSIITFQEPICILSDKGRNVLLVILLTGALCFEIIAGVFPS 289
```

RESULT 9	
Q9KEG8	
ID Q9KEG8	PRELIMINARY; PRT; 401 AA

AC Q9KEG8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein BH0884.
GN BH0884.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001510; BAB04603.1; -.
DR PIR; D83760; D83760.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS50850; MFS; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 401 AA; 44284 MW; E7386A8BBE1DCF016 CRC64;

	Query Match	7.3%;	Score 121.5;	DB 16;	Length 401;
	Best Local Similarity	24.4%;	Pred. No. 0.084;		
	Matches	73;	Conservative	45;	Mismatches 94; Indels 87; Gaps 19;
QY	3 LMNGVL---PFYPQPRHAAGFSVPELLIVILVELALAAFFLLIPGIRGRSHRWFWLYRVLL	59			
Dd	56 LYIGILLASPFIEKPLRRYGK-PITTIGLMAMIVS---LVLVP-LMQAFWFVFV---L	106			
QY	60 SLFTGAELIVAHFSAEMFVGTVN-----TNTSYKAFAARVTARVGLLVGLEGINITLT	113			
Dd	107 RMIIGIADNMVHFATQVWITTSPMKKRGNISFYGFA-----FGLRGFLGPFM----	155			
QY	114 GTPVHQLNETIDY-----NEQFTW----RLKENYAABYANALEKGLPDPVLYLAKEPTPS	164			
Dd	156 -TRLLOINEEFLPFIISAITSFAAMULLIRLRNRYPAQ--DYETG-----	196			
QY	165 SPCGLHYOHLAGHYASATLMVAFCFWLLSNVLISTPAPIYG--GLAL---LTTGGAFA	217			
Dd	197 SOSGTWRYYKAVVKLGWFALLPAFCYGYLESSLHG-N-FPVYGLRSGLTVEQVSILLEPAFV	255			
QY	218 LFGVFALASISSVPLCPPL-RLGSSALTTOYGAAFWTLATGVLCFLGGAV-VSILOYV	273			
Dd	256 VGGI-----ITOMPLFLGSDRMGRKPL-----IMVLFFGGATFMSTMYLL	295			

```

RESULT 10
Q82128
ID Q82128 PRELIMINARY; PRT; 467 AA.
AC Q82128;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative membrane protein.
GN STY4802 OR T4498.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;

```

RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Cornerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagsels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant *Salmonella*
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; Pubmed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
DR EMBL; AL627283; CAD06924.1; -.
DR EMBL; AE016849; AAC071945.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0015556; F:C4-dicarboxylate transporter activity; IEA.
DR GO; GO:0015740; P:C4-dicarboxylate transport; IEA.
DR InterPro; IPR004669; Dcuc.
DR Pfam; PF03606; Dcuc; 1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 467 AA; 49844 MW; 5CF0D6293B252926 CRC64;

	Query Match	6.9%; Score 115; DB 16; Length 467;
	Best Local Similarity	24.2%; Pred. NO. 0.35;
	Matches	63; Conservative 28; Mismatches 75; Indels 94; Gaps 11;
QY	27 VILVFLALAAFLLLPGIRGHSRWFVLVRLSLFPGAELVAVHESAEWFGVTNTS	86
Db	264 VLLLF--AASFVMIVYGAVLGWM-----AESGVFLAAIIVGI-TRMG	307
QY	87 YKAFSA----ARVTARVGLVGL-EGI---NITLTGPVHQINETIDYNEGEFTWRLK	135
Db	308 EEAFTSTPIDGARDLLGVALIIIGARGIVVMDNGMTHIIHS-----	351
QY	136 ENYAAEYANALEKGLPDPVLYLAEKFTPESSPCGLYHQYHLAGHYASATLMVAFCEWLLSN	195
Db	352 -----AESLVSGL-----STTIFINVYWL-E	372
QY	196 VLLSTPAPLYGGALLTTGAFAFLFGVFALASISSVPICPLRLGSSALLTQYGAAFWTTLA	255
Db	373 VLLSLVLPSSSGLAVALTMPIMAPLADFAHVQRDLV-----VTAYQSASGIYNLV	421
QY	256 TGVLCFLGG-AVVSLOQYR 274	
Db	422 TPTSAVVMGGLAIRVPIYR 441	

RESULT	11		
Q9ZBJ1			
ID	Q9ZBJ1	PRELIMINARY;	PRT; 815 AA.
AC	Q9ZBJ1;		
DT	01-MAY-1999	(TREMBLrel. 10,	Created)
DT	01-MAY-1999	(TREMBLrel. 10,	Last sequence update)
DT	01-OCT-2003	(TREMBLrel. 25,	Last annotation update)
DE	Putative efflux protein.		
GN	SCO6483 OR SC9C7.19.		
OS	Streptomyces coelicolor.		
OC	Bacteria;		
OC	Actinobacteridae;		
OC	Streptomycinae;		
OC	Streptomycetaceae;		
OC	Streptomyces.		
OX	NCBI_TaxID=1902;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=A3(2);		
RA	Seeger K.J., Harris D.;		
RL	Submitted (JAN-1999)		
	to the EMBL/GenBank/DBJ		
	databases.		

[2]
RN SEQUENCE FROM N.A.
RP STRAIN=A3(2);
RC Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.,
RA Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome."; Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=2196410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2)."; Nature 417:141-147(2002).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AL939128; CAA22731.1; -.
DR PIR: T35970; T35970.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0015520; F:tetracycline:hydrogen antiporter activity; IEA.
DR GO: GO:0005215; F:transporter activity; IEA.
DR GO: GO:0015904; P:tetracycline transport; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR008969; Carboxypept_reg.
DR InterPro: IPR008454; Cna_B.
DR InterPro: IPR004638; Efflux_EmbB.
DR InterPro: IPR007114; MFS.
DR InterPro: IPR005828; Sub_transporter.
DR InterPro: IPR001958; TCR_TetA.
DR InterPro: IPR001411; TCR_TetB.
DR Pfam: PF05738; Cna_B_2.
DR Pfam: PF00083; sugar tr; 1.
DR PRINTS: PR01035; TCRTERA.
DR PRINTS: PR01036; TCRTERB.
DR TIGRFAMs: TIGR00711; efflux_EmbB; 1.
DR PROSITE: PS00850; MFS; 1.
KM Transmembrane; Complete proteome.
SQ SEQUENCE 815 AA; 84282 MW; 9AD3C72DF086DB96 CRC64;

Query Match 6.9%; Score 115; DB 16; Length 815;
Best Local Similarity 24.2%; Pred. No. 0.69;
Matches 85; Conservative 25; Mismatches 107; Indels 134; Gaps 15;

QY 12 PQPRHAAGFSVPLILVFLALAAFLILPGIRGSRWFVLVRVLSLFIGAE----- 66
DB 226 PRPR-----VKPRLDILGALLLAASTCLVLTLSWGTEYAWGSVILGLAAGAVAAATLL 280

QY 67 -IVAVHFSAEWFGVTNTNTSYKAF--SAARVTARVGLVGLGEGINITLTGTPVHQINET 123
DB 281 FLVAEHFAPEPLI-----PLRLFRDSIFNVTALVGLVVG-----VALFG----- 319

QY 124 IDVNEQFTWRLKENYAAY-----ANALEKGLPDVLYLAKEKTPS----- 164
DB 320 -----AASYLPTFLQWVDGASATESGL-----IMLPMMGVVGASI 355

QY 165 -----SPGGLHYOYHLAGHYASATLWAFCFWLLSNVLSTPAPLYG-GLALLTTGAF 217
DB 356 ISGQLISRTGHRIRHPILGSLSV-----GMWLLSRDLADTSRLHYISIMAAVLGAGIGL 410

QY 218 LFGVFAIASISSVPLCPRLRGSSALLTQY-----GAAFWTLATGVLCFLGAVVS 269
DB 411 VMPVLVAVQNSVR--FTDLGTATSNANFYFRQIGSGVGAAVFGTLFAG----- 456

QY 270 LQYVRPALARLLDQSAKDCSQERGSPILGDPHLKQALPDKCITTNL 320
DB 457 -----RLTDALADRIPEAG-----VGLPDAEAITPOL 484

RESULT 12

QY 0974G1 PRELIMINARY; PRT; 446 AA.
AC 0974G1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Putative transporter.
GN ST0691.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Anai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermophilic
RL Crenarchaeon, Sulfolobus tokodaii strain7.";
DR DNA Res. 8:123-140(2001).
DR EMBL: AP000983; BAB65697.1; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005215; F:transporter activity; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR007114; MFS.
DR InterPro: IPR005828; Sub_transporter.
DR InterPro: IPR005829; Sug_transporter.
DR Pfam: PF00083; sugar tr; 1.
DR PROSITE: PS00850; MFS; 1.
DR PROSITE: PS00217; SUGAR_TRANSPORT 2; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 446 AA; 48375 MW; F4495E31897E06CB CRC64;

Query Match 6.9%; Score 114.5; DB 17; Length 446;
Best Local Similarity 19.8%; Pred. No. 0.36;
Matches 65; Conservative 52; Mismatches 90; Indels 121; Gaps 12;

QY 6 GVLPRYQPRHAAGFSVPLILVFLALAAFLILPGIRGSRWFVLVRVLSLFIGA 65
DB 63 GILSDFKRRKRMYGIDVTLMSIGALQAISQNYEL-----FVSRLLGMGIGA 111

QY 66 E-----IVAVHFSAEWFGTV--NTNTSYKAFSAARVTARVGLLV 103
DB 112 DYVLSPIIVAENAEAKRGKLMVTFPAVLMGLGAVFAFVVDQISSIFLPSSLVWRVVLGV 171

QY 104 G-LEGINITLTGTPVHQINETIDY-----NEQFTWRLKENYA-----AYANALE 147
DB 172 GAIPATISIMAR--RKLTEITLQFLTKVPDENLQIKITNYGLLGINVDEKEFINRLK 228

QY 148 KGLPDVLYLAKEKTPSPGGLHYOYHLAGHYASATLWAFCFWLLSNVLSTPAPLYGG 207
DB 229 ASLP-----FIIVASVLMLLDYISST----- 250

QY 208 LALLTTGAFALFGVFAIASISSVPLCPRLRGSSALLTQYGAAFWTLATGVCLFL----- 263
DB 251 -----FAIYGPVIVAS-----NLGLPTITFTTYVAQFPAGIPGQLCIYLDKV 293

QY 264 -----GGAVVSLQYRPSALRTLLD 283
DB 294 GRKILITIGYAGVALMLVMSLL--LLD 319

RESULT 13

Q82LV7
ID Q82LV7 PRELIMINARY; PRT; 792 AA.
AC Q82LV7;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative efflux protein.
GN SAV1903.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RT Nat. Biotechnol. 21:526-531(2003).
RL EMBL; AP005028; BAC69614.1;
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0015520; F:tetracycline:hydrogen antiporter activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0015904; P:tetracycline transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR008969; Carboxypept_reg.
DR InterPro; IPR004638; Efflux_EmrB.
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub transporter.
DR InterPro; IPR001411; TCR_TetB.
DR Pfam; PF00083; sugar_cr; 1.
DR PRINTS; PRO1036; TCRTEB.
DR TIGRFAMs; TIGR00711; efflux_EmrB; 1.
DR PROSITE; PS50850; MFS; 1.
KM Complete proteome.
SQ SEQUENCE 792 AA; 82256 MW; EEF543D0297D9EA1 CRC64;

Query Match

Best Local Similarity 6.8%; Score 113.5; DB 16; Length 792;
Matches 79; Conservative 31; Mismatches 111; Indels 105; Gaps 15;

QY 25 LIVILVFLAASFLILPGIRGSRWFLVRLSLFIGAEIVAV-----HSAEMFV 78
DB 221 LDVLGALLAVASTCAVLLTSGTEYADSRITILGAGACVAAVLFLVEHFAPEPLI 280
QY 79 GTVNTNTSYKAF--SAARVTARVGLVLEGINITLTGTVPVHQLNETIDYNEQFTWRLKE 136
DB 281 -----PLRLFRDSVFENVTLGLVLTG-----VALFG----- 306
QY 137 NYAAEY-----ANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYASAT--- 183
DB 307 --AASYLPTFLQWVDGASATSEGL---LML-----PMAAGIVGASIIAGQLISHTGRY 354
QY 184 -LW-----VAFCFWLLSNVLSTPAPLYG-GLALLTTGAFALFGVAFALASISSVPLCP 234

DB 355 RMWPIIGSVLAAGVMWLLSRLEADTPRLQYSIMWAVLGAGIGMVPVLVAVQNSVR--P 412
QY 235 LRLGSSALTTOYGAAFWVTLATVLCFLGGAVVSLQYRPSALRTLLDQSAKDCSÖERG 294
DB 413 ADLGTATISANNYFRQ-----IGSVGA-----AVFGTLFARLADALEER- 452
QY 295 GSPLILDDPLHKQALPDLCITTNL 320
DB 453 -----LPTRAGARLPDPESTLPQL 471

RESULT 14

Q82K36
ID Q82K36 PRELIMINARY; PRT; 467 AA.
AC Q82K36;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Putative arginine repressor.
GN STM4464.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RT Nature 413:852-856(2001).
RL EMBL; AE008909; AAU23283.1;
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0015556; F:C4-dicarboxylate transporter activity; IEA.
DR GO; GO:0015740; P:C4-dicarboxylate transport; IEA.
DR InterPro; IPR004669; DcuC.
DR Pfam; PF03606; DcuC; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 467 AA; 49890 MW; 0E60DE1B5F0475FE CRC64;

Query Match

Best Local Similarity 6.8%; Score 113; DB 16; Length 467;
Matches 62; Conservative 29; Mismatches 75; Indels 94; Gaps 11;

QY 27 VILVFLAASFLILPGIRGSRWFLVRLSLFIGAEIVAVHSAEMFVGTNTNTS 86
DB 264 VLMIF--AASFVAVMITYGAVLGWMM-----AEISGVFLAAAIIVGVI-TRMG 307
QY 87 YKAFSA-----ARVTARVGLVGL-EGI-----NITLTGTVPVHQLNETIDYNEQFTWRLK 135
DB 308 EBAFTSTFIDGARDLGLVALITIGARGIVVMDNGMITTHLS----- 351
QY 136 ENVAEYANALEKGLPDPVLYLAKEFTPSSPCGLYHQYHLAGHYASATLWVAFCFWLLSN 195
DB 352 -----AESLVSGL-----STIFINVTYML--E 372
QY 196 VLLSTPAPLYGGLALLITGAFALFGVAFALASISSVPLCPRLGSSALTTOYGAAFWVTIA 255
DB 373 VLLSFLVPSSSGLAVLTWPIAPLADFAHQRDLY-----VTAYQASAGIVNLV 421
QY 256 TGVLCFLGG-AVVSLOQYR 274
DB 422 TPTSAYVMGGLAIARVPYR 441

RESULT 15

047547

ID	047547	PRELIMINARY;	PRT;	509 AA.
AC	047547;			
DT	01-JUN-1998	(TREMBLrel. 06, Created)		
DT	01-JUN-1998	(TREMBLrel. 06, Last sequence update)		
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)		
DE	Cytochrome c oxidase subunit I (EC 1.9.3.1) (Cytochrome c oxidase polypeptide I).			
DE	COX1.			
OS	Chlamydomonas eugametos.			
OG	Mitochondrion.			
OC	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;			
OX	Chlamydomonadaceae; Chlamydomonas.			
OX	NCBI_TaxID=3053;			
RN	[1]			
RE	SEQUENCE FROM N.A.			
RX	MEDLINE=98145434; PubMed=9484440;			
RA	Denovan-Wright E.M., Nedelcu A.M., Lee R.W.;			
RT	"Complete sequence of the mitochondrial DNA of Chlamydomonas eugametos.";			
RL	Plant Mol. Biol. 36:285-295(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Denovan-Wright E.M., Nedelcu A.M., Lee R.W.;			
RL	Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.			
CC	-1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B (BY SIMILARITY).			
CC	-1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME C + 2 H(2)O.			
CC	-1- PATHWAY: RESPIRATORY CHAIN, TERMINAL STEP.			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.			
DR	EMBL; AF008237; AAC9342.1; -			
DR	PIR; T11043; T11043.			
DR	HSSP; P98002; IAR1.			
DR	GO; GO:0019866; C:inner membrane; IEA.			
DR	GO; GO:0016021; C:integral to membrane; IEA.			
DR	GO; GO:0005746; C:mitochondrial electron transport chain; IEA.			
DR	GO; GO:0005739; C:mitochondrion; IEA.			
DR	GO; GO:0009481; F:aa3-type cytochrome c oxidase; IEA.			
DR	GO; GO:0009482; F:ba3-type cytochrome c oxidase; IEA.			
DR	GO; GO:0009483; F:ca3-type cytochrome c oxidase; IEA.			
DR	GO; GO:0009485; F:cbb3-type cytochrome c oxidase; IEA.			
DR	GO; GO:0004129; F:cytochrome-c oxidase activity; IEA.			
DR	GO; GO:0016491; F:oxidoreductase activity; IEA.			
DR	GO; GO:0006118; P:electron transport; IEA.			
DR	GO; GO:0006810; P:transport; IEA.			
DR	InterPro; IPR000883; COX1.			
DR	Pfam; PF00115; COX1; 1.			
DR	PRINTS; PR01165; CYCOXIDASEI.			
DR	PROSITE; PS00077; COX1; 1.			
KM	Copper; Electron transport; Heme; Inner membrane; Membrane; Oxidoreductase; Respiratory chain; Transmembrane; Transport; Mitochondrion.			
KM	SEQUENCE 509 AA; 55847 MW; 8EAA06905EA00E12 CRC64;			

Query Match	6.8%;	Score 113;	DB 8;	Length 509;
Best Local Similarity	21.2%;	Pred. No. 0.56;		
Matches 66;	Conservative 38;	Mismatches 118;	Indels 90;	Gaps 14;

```
QY      26 IVLVFLALLAASFLLLLEPGIRGHSR-----FWLVRVLLSLFICA 65
          |||::|||
Db      64 IIMLLFMVPALF-----GGFGNWLVLPIIGADMAFPRLNNISFWLNPSALGLILLSS 116
          :|::|
QY      66 EIVAVHFSAEWFFVGVTNTNTSYKAFSAARVTARVGLV-----GLE---GINITLTGT 115
          ::|||:
Db     117 TMVEOGAGTGW-----TAYPPLSIOSTGAAYDLAILSLHLNGLSSTIGSINITLVITA 168
```

[illegible]

```
Search completed: February 23, 2004, 19:19:33
Job time : 83 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 21:58:00 / Search time 1965.03 Seconds

(without alignments)
10984.465 Million cell updates/sec

Title: US-09-936-456-3

Perfect score: 498
Sequence: 1 cctagcgtgcgcgtctgcc.....aaaaaaataaagtcgacc 498

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenBml:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	498	100.0	498	6 AX035348	AX035348 Sequence
2	425	85.3	1441	6 AX035346	AX035346 Sequence
3	392	78.7	2684	6 AX714511	AX714511 Sequence
4	392	78.7	2684	6 AK056896	AK056896 Homo sapi
5	392	78.7	156534	2 AC009700	AC009700 Homo sapi
6	392	78.7	166937	2 AC012255	AC012255 Homo sapi
7	392	78.7	171444	2 AC087790	AC087790 Homo sapi
8	392	78.7	181312	9 AC091117	AC091117 Homo sapi
9	374	75.1	1740	9 HSM806241	BX537581 Homo sapi
10	217	43.6	217	9 HSI83H12F	Z57552 H. sapiens C
11	166	33.3	1923	9 BC029819	BC029819 Homo sapi
12	136	27.3	218	9 HSI83A12R	Z59954 H. sapiens C
13	26	5.2	214866	9 AC069544	AC069544 Homo sapi
14	26	5.2	254436	3 AB014827	AB014827 Plasmodiu
15	25	5.0	55790	8 AB025602	AB025602 Arabidops
16	24	4.8	95824	8 U90439	U90439 Arabidopsis
17	24	4.8	115811	9 AL136377	AL136377 Human DNA
18	24	4.8	157965	2 AC011220	AC011220 Homo sapi
19	24	4.8	186184	2 AC140136	AC140136 Homo sapi
20	24	4.8	188769	2 AP002809	AP002809 Homo sapi
21	24	4.8	204267	9 AC010736	AC010736 Homo sapi
22	23	4.6	90157	9 AL355552	AL355552 Human DNA
23	23	4.6	91851	8 ATT16124	AL138659 Arabidops
24	23	4.6	125623	3 AC115599	AC115599 Dictyoste
25	23	4.6	137764	2 AC009655	AC009655 Homo sapi
26	23	4.6	153569	5 BX649231	BX649231 Zebrafish
27	23	4.6	153788	2 AC027472	AC027472 Homo sapi
28	23	4.6	154677	2 AC016196	AC016196 Homo sapi
29	23	4.6	154677	2 AC016196	AC016196 Homo sapi
30	23	4.6	157326	2 AC011024	AC036148 Homo sapi
31	23	4.6	161079	2 AC036148	AC090980 Homo sapi
32	23	4.6	182881	9 AC090980	AC068611 Homo sapi
33	23	4.6	183016	2 AC145790	AC145790 Silurana
34	23	4.6	193665	9 AC105314	AC105314 Homo sapi
35	23	4.6	193929	9 AC090797	AC090797 Homo sapi
36	23	4.6	194122	2 AC138778	AC138778 Homo sapi
37	23	4.6	194122	2 AC097902	AC097902 Rattus no
38	23	4.6	250029	9 AE014310	AE014310 Homo sapi
39	22	4.4	6415	6 AX251125	AX251125 Sequence
40	22	4.4	43900	9 AC005265	AC005265 Homo sapi
41	22	4.4	99687	9 AL603790	AL603790 Human DNA
42	22	4.4	108964	8 AC073391	AC073391 Oryza sat
43	22	4.4	110000	1 AP006628	Continuation (6 of
44	22	4.4	110250	9 AC010104	AC010104 Homo sapi
45	22	4.4	110250	9 AC010104	AC010104 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AX035348 498 bp DNA linear PAT 15-NOV-2000
DEFINITION Sequence 3 from Patent WO0053748.
ACCESSION AX035348
VERSION AX035348.1 GI:11191065
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Bruck,C.E., Vinals,Y.D., Coche,T. and Cassart,J.P.
AUTHORS Novel compounds
TITLE Patent: WO 0053748-A 3 14-SEP-2000;
JOURNAL

BRUCK CLAUDE ELVIRE MARIE (BE) ; SMITHKLINE BEECHAM BIOLOG (BE) ;
VINALS Y DE BASSOLS CARLOTA (BE) ; COCHE THIERRY (BE) ; CASSART
JEAN POL (BE)

FEATURES
source

Location/Qualifiers
1..498
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 498; DB 6; Length 498;
Best Local Similarity 100.0%; Pred. No. 6.2e-269;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCTAGCGTGGCGCTCTGCGCGCTCCCGCTAGGCTCTCCGCGCTCACCCTCAGTACG 60
DB 1 CTCTAGCGTGGCGCTCTGCGCGCTCCCGCTAGGCTCTCCGCGCTCACCCTCAGTACG 60
QY 61 AGCGCGCTCTTGGGTACAGCTGGCAACCGCGCTCTGTGCTCTTCTCGAGGGGCC 120
DB 61 AGCGCGCTCTTGGGTACAGCTGGCAACCGCGCTCTGTGCTCTTCTCGAGGGGCC 120
QY 121 GTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTCCGACCCCTTCTGACCAAGCGCC 180
DB 121 GTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTCCGACCCCTTCTGACCAAGCGCC 180
QY 181 AAGGACTGCAGCCAGAGAGAGGGGGCTCACTCTTATCTCTCGGCAACCCACTGCACAAG 240
DB 181 AAGGACTGCAGCCAGAGAGAGGGGGCTCACTCTTATCTCTCGGCAACCCACTGCACAAG 240
QY 241 CAGGCGCTCTCCAGACTTAATGTATCACCACCTAACCCTGTGAGGGGGACCCATCTG 300
DB 241 CAGGCGCTCTCCAGACTTAATGTATCACCACCTAACCCTGTGAGGGGGACCCATCTG 300
QY 301 GACTCCTTCCCGCCTTGGGACATCGAGCGCCGGAGACAGTGCCCGCAGCGCTGGGCC 360
DB 301 GACTCCTTCCCGCCTTGGGACATCGAGCGCCGGAGACAGTGCCCGCAGCGCTGGGCC 360
QY 361 AGGAGAGCTCCAGAGAGGCACTGAGCGCTGTGCGGAGGCTCGGACATCGCAGGC 420
DB 361 AGGAGAGCTCCAGAGAGGCACTGAGCGCTGTGCGGAGGCTCGGACATCGCAGGC 420
QY 421 ACCAGGAGAAAGTCTCCGCGCGCATCTGTAAATAAACCTTTTCTTTGTTTAA 480
DB 421 ACCAGGAGAAAGTCTCCGCGCGCATCTGTAAATAAACCTTTTCTTTGTTTAA 480
QY 481 AAAAAATAAAAAGTCGACC 498
DB 481 AAAAAATAAAAAGTCGACC 498

RESULT 2

AX035346 1441 bp DNA linear PAT 15-NOV-2000
LOCUS AX035346
DEFINITION Sequence 1 from Patent WO0053748.
ACCESSION AX035346
VERSION AX035346.1 GI:11191064
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL

1 Bruck, C.E., Vinals, Y.D., Coche, T. and Cassart, J.P.
Novel compounds
Patent: WO 0053748-A 1 14-SEP-2000;
BRUCK CLAUDE ELVIRE MARIE (BE) ; SMITHKLINE BEECHAM BIOLOG (BE) ;
VINALS Y DE BASSOLS CARLOTA (BE) ; COCHE THIERRY (BE) ; CASSART
JEAN POL (BE)
Location/Qualifiers
1..1441
/organism="Homo sapiens"
/mol_type="unassigned DNA"

FEATURES
source

1..1441
/organism="Homo sapiens"
/mol_type="unassigned DNA"

/db_xref="taxon:9606"

ORIGIN

Query Match 85.3%; Score 425; DB 6; Length 1441;
Best Local Similarity 100.0%; Pred. No. 9.9e-228;
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 GCGCGCGCTTGTGGGTACAGCTGGCAACCGCGCTCTGTGCTCTTCTCGAGGGGCCG 121
DB 998 GCGCGCGCTTGTGGGTACAGCTGGCAACCGCGCTCTGTGCTCTTCTCGAGGGGCCG 1057
QY 122 TGGTGAAGTCTCCAGTATGTTCCGCCAGCGCTCTTCCGACCCCTTCTGACCAAGCGCCA 181
DB 1058 TGGTGAAGTCTCCAGTATGTTCCGCCAGCGCTCTTCCGACCCCTTCTGACCAAGCGCCA 1117
QY 182 AGGACTGCAGCCAGAGAGAGGGGGCTCACTCTTATCTCTCGGCAACCCACTGCACAAGC 241
DB 1118 AGGACTGCAGCCAGAGAGAGGGGGCTCACTCTTATCTCTCGGCAACCCACTGCACAAGC 1177
QY 242 AGCGCGCTCTCCAGACTTAATGTATCAACCACTAACCCTGTGAGGGGGACCCATCTGG 301
DB 1178 AGCGCGCTCTCCAGACTTAATGTATCAACCACTAACCCTGTGAGGGGGACCCATCTGG 1237
QY 302 ACTCCTTCCCGCCTTGGGACATCGAGCGCGGAGAGAGTGGCCCGCAGCGCTGGGCCA 361
DB 1238 ACTCCTTCCCGCCTTGGGACATCGAGCGCGGAGAGAGTGGCCCGCAGCGCTGGGCCA 1297
QY 362 GGAAGCTCCAGAGAGGCACTGAGCGCTGCTGGCGAGGCTCGGACATCGCAGGCA 421
DB 1298 GGAAGCTCCAGAGAGGCACTGAGCGCTGCTGGCGAGGCTCGGACATCGCAGGCA 1357
QY 422 CCAGGAGAAAGTCTCTGGGGGATCTGTAAATAAACCTTTTCTTTGTTTAA 481
DB 1358 CCAGGAGAAAGTCTCTGGGGGATCTGTAAATAAACCTTTTCTTTGTTTAA 1417
QY 482 AAAAA 486
DB 1418 AAAAA 1422

RESULT 3

AX714511/c 2684 bp DNA linear PAT 15-APR-2003
LOCUS AX714511
DEFINITION Sequence 1195 from Patent EP1293569.
ACCESSION AX714511
VERSION AX714511.1 GI:29889464
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
AUTHORS

1 Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,
Yamamoto, J.I., Isono, Y., Hio, Y., Otsuka, K., Nagai, R.,
Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and
Masuko, Y.
Full-length cDNAs
Patent: EP 1293569-A 1195 19-MAR-2003;
Helix Research Institute (JP) ; Research Association for
Biotechnology (JP)
Location/Qualifiers
1..2684
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

FEATURES
source

1..2684
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 78.7%; Score 392; DB 6; Length 2684;
Best Local Similarity 100.0%; Pred. No. 4.2e-209;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GCGCTCTGTGCTCTTCTCTCGAGGGGGCGGTGAGTCTCCAGTATGTTGGCCGACG 150
|||||

Db 2343 GGGCTCCTGTGCTCTCTCTCGAGAGGGCCGTGTGAGTCTCCAGTATGTGGCCAGC 2284

QY 151 GCTCTTGGACCCCTTGTGACCAAGGCCAAGGACTGACGACGAGAGAGAGGGGGCTCA 210

Db 2283 GCTCTTGGACCCCTTGTGACCAAGGCCAAGGACTGACGACGAGAGAGAGGGGGCTCA 2224

QY 211 CCTCTTATCCTCGGCGACCCACTGACAGAGAGGGCCCTCTCCAGACTTAAATGTATC 270

Db 2223 CCTCTTATCCTCGGCGACCCACTGACAGAGAGGGCCCTCTCCAGACTTAAATGTATC 2164

QY 271 ACCACTAAGCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGAGACATCGCAGG 330

Db 2163 ACCACTAAGCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGAGACATCGCAGG 2104

QY 331 CCGGGAAGAGTGGCCCGGACGCTGGCCAGAGAGCTCCAGAGAGGGGACATGAGCGCT 390

Db 2103 CCGGGAAGAGTGGCCCGGACGCTGGCCAGAGAGCTCCAGAGAGGGGACATGAGCGCT 2044

QY 391 GCTGGCGGAGGCTCGGACATCGGACAGGACCAAGGAAAGTCTCTGGGGCGATCTGTA 450

Db 2043 GCTGGCGGAGGCTCGGACATCGGACAGGACCAAGGAAAGTCTCTGGGGCGATCTGTA 1984

QY 451 AATAAACCTTTTCTTTCTTTGTTTAAAAA 482

Db 1983 AATAAACCTTTTCTTTCTTTGTTTAAAAA 1952

RESULT 4

AK056896/c 2684 bp mRNA linear PRI 01-AUG-2002

LOCUS AK056896 Homo sapiens cDNA FLJ32334 f1s, clone PROST2005426.

ACCESSION AK056896

VERSION AK056896.1 GI:16552419

KEYWORDS oligo capping; f1s (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE

AUTHORS

1 Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S., Ono,Y., Houta,T., Hiraoka,S., Murakawa,K., Takiguchi,S., Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M., Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuna,M., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahara,K., Masuno,Y., Nagai,K. and Isogai,T.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

2 (bases 1 to 2684)

Isogai,T., Otsuki,T. and Sugiyama,T.

Direct Submission

Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI); cDNA library Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

COMMENT

FEATURES

source

1. 2684

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="PROST2005426"

/tissue_type="prostate"

/clone_lib="PROST2"

/note="Cloning vector: PME18SFL3"

CDS

56. .952

/note="unnamed protein product"

/codon_start=1

/protein_id="BAB71304.1"

/db_xref="GI:16552420"

/translation="MATLGHTRPFYAGPKPFPMDTTIASIMIFLTALATFIVLPGRKTRFLRLRVTSIFIGAILGTPOQNETINYNENFTWRLGENYAEYAKALEKGLDPVLYLAEXTPRSPCGIXRQYRLAGHYTSAMLWVFLCWLANWLSMPLVYGGVMLATGIFOLALLPFSMASTLSPCLHLGASVLTHTGPAFWITVITGLCVL LGILAMVAHRMQPHRLKAFNQSVDDEPMLMSPEEGLLSPRYRSMADBPXSQDIPL SEASSTKAYCKEAPKPDPCAL"

Query Match 78.7%; Score 392; DB 9; Length 2684;

Best Local Similarity 100.0%; Pred. No. 4.2e-209;

Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GGGCTCCTGTGCTCTTCTCGAGAGGGCCGTGTGAGTCTCCAGTATGTTCCGCCAGC 150

Db 2343 GGGCTCCTGTGCTCTTCTCGAGAGGGCCGTGTGAGTCTCCAGTATGTTCCGCCAGC 2284

QY 151 GCTCTTGGACCCCTTGTGACCAAGGCCAAGGACTGACGACGAGAGAGGGGGCTCA 210

Db 2283 GCTCTTGGACCCCTTGTGACCAAGGCCAAGGACTGACGACGAGAGAGGGGGCTCA 2224

QY 211 CCTCTTATCCTCGGCGACCCACTGACAGAGAGGGCCCTCTCCAGACTTAAATGTATC 270

Db 2223 CCTCTTATCCTCGGCGACCCACTGACAGAGAGGGCCCTCTCCAGACTTAAATGTATC 2164

QY 271 ACCACTAAGCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGAGACATCGCAGG 330

Db 2163 ACCACTAAGCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGGAGACATCGCAGG 2104

QY 331 CCGGGAAGAGTGGCCCGGACGCTGGCCAGAGAGCTCCAGAGAGGGGACATGAGCGCT 390

Db 2103 CCGGGAAGAGTGGCCCGGACGCTGGCCAGAGAGCTCCAGAGAGGGGACATGAGCGCT 2044

QY 391 GCTGGCGGAGGCTCGGACATCGGACAGGACCAAGGAAAGTCTCTGGGGCGATCTGTA 450

Db 2043 GCTGGCGGAGGCTCGGACATCGGACAGGACCAAGGAAAGTCTCTGGGGCGATCTGTA 1984

QY 451 AATAAACCTTTTCTTTCTTTGTTTAAAAA 482

Db 1983 AATAAACCTTTTCTTTCTTTGTTTAAAAA 1952

RESULT 5

AC009700 156534 bp DNA linear HTG 20-APR-2000

LOCUS AC009700 Homo sapiens chromosome 15 clone RP11-163P10 map 15, WORKING DRAFT

DEFINITION

SEQUENCE, 15 unordered pieces.

ACCESSION AC009700

VERSION AC009700.4 GI:7622346

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 156534)

Bitren,B., Linton,L., Nusbbaum,C., Lander,E., Allen,N., Anderson,M., Baker,U., Baldwin,J., Barra,N., Beckerly,R., Benn,J., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G., Hagos,B., Heaford,A., Horton,C., Howland,J.C., Jones,C., Kann,L., Karatas,A., Lehoczkj,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGuirk,A., McKernan,K., McLaughlin,J., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,

TITLE
JOURNAL
COMMENT

Naylor J., Nioff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
Peterson, K., Polara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.
Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 20, 2000 this sequence version replaced gi:6056272.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L2336
Center clone name: 163_P.10

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 108837 bases at least Q40
Consensus quality: 130230 bases at least Q30
Consensus quality: 145227 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 155134; sum-of-contigs
Quality coverage: 3.2 in Q20 bases; agarose-fp
Quality coverage: 3.3 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 1169: contig of 1169 bp in length
1170 1269: gap of 100 bp
1270 3335: contig of 2066 bp in length
3336 3435: gap of 100 bp
3436 5698: contig of 2263 bp in length
5699 5798: gap of 100 bp
5799 9216: contig of 3418 bp in length
9217 9316: gap of 100 bp
9317 15633: contig of 6317 bp in length
15634 15733: gap of 100 bp
15734 21746: contig of 6013 bp in length
21747 21846: gap of 100 bp
21847 30351: contig of 8505 bp in length
30352 30451: gap of 100 bp
30452 36509: contig of 6058 bp in length
36510 36609: gap of 100 bp
36610 45279: contig of 8670 bp in length
45280 45379: gap of 100 bp
45380 56851: contig of 11472 bp in length
56852 56951: gap of 100 bp
56952 67374: contig of 10423 bp in length
67375 67474: gap of 100 bp
67475 79534: contig of 12060 bp in length
79535 79634: gap of 100 bp
79635 95432: contig of 15798 bp in length
95433 95532: gap of 100 bp
95533 120801: contig of 25269 bp in length
120802 120901: gap of 100 bp
120902 156534: contig of 35633 bp in length.
```

FEATURES

source

1.156534
/organism="Homo sapiens"
/mol_type="genomic DNA"

```
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-163P10"
/clone_lib="RPC1-11 Human Male BAC"
1.1169
/note="assembly_fragment"
1270.3335
/note="assembly_fragment"
3436.5698
/note="assembly_fragment"
5799.9216
/note="assembly_fragment"
9317.15633
/note="assembly_fragment"
15734.21746
/note="assembly_fragment"
21847.30351
/note="assembly_fragment"
30452.36509
/note="assembly_fragment"
30452.36509
/note="assembly_fragment"
vector_side:left"
36610.45279
/note="assembly_fragment"
45380.56851
/note="assembly_fragment"
56952.67374
/note="assembly_fragment"
67475.79534
/note="assembly_fragment"
79635.95432
/note="assembly_fragment"
95533.120801
/note="assembly_fragment"
120902.156534
/note="assembly_fragment"
```

ORIGIN

```
Query Match 78.7%; Score 392; DB 2; Length 156534;
Best Local Similarity 100.0%; Pred. No. 6e-209;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GGGCTCCTGTGCTCTTCTCGAGAGGCGCTGTGAGTCTCCAGTATGTTCCGCCAGC 150
Db 96131 GGGCTCCTGTGCTCTTCTCGAGAGGCGCTGTGAGTCTCCAGTATGTTCCGCCAGC 96190

QY 151 GCTCTTGCACCTTCTGTGACCAAGCGCAAGAGTGTGAGCCAGAGAGGGGCTCA 210
Db 96191 GCTCTTGCACCTTCTGTGACCAAGCGCAAGAGTGTGAGCCAGAGAGGGGCTCA 96250

QY 211 CCTCTTATCCTGCGGAGACCACTGTGACATCTCTCCCGCTTGGGACATCGCAGG 270
Db 96251 CCTCTTATCCTGCGGAGACCACTGTGACATCTCTCCCGCTTGGGACATCGCAGG 96310

QY 271 ACCACTAACCTGTGAGGGGAGCAATCTGACTCTTCCCGCTTGGGACATCGCAGG 330
Db 96311 ACCACTAACCTGTGAGGGGAGCAATCTGACTCTTCCCGCTTGGGACATCGCAGG 96370

QY 331 CCGGGAAGCAGTGGCCCGCAGGCTTGGCCAGAGAGCTTCCAGGAAGGCACTGAGCGCT 390
Db 96371 CCGGGAAGCAGTGGCCCGCAGGCTTGGCCAGAGAGCTTCCAGGAAGGCACTGAGCGCT 96430

QY 391 GCTGGCGGAGGCTCGGACATCCGAGGACCAAGGAAGTCTCTGGGGCGATCTGTA 450
Db 96431 GCTGGCGGAGGCTCGGACATCCGAGGACCAAGGAAGTCTCTGGGGCGATCTGTA 96490

QY 451 AATAAACCTTTTCTTTGTTTAAATA 482
Db 96491 AATAAACCTTTTCTTTGTTTAAATA 96522
```

RESULT 6

AC012255
LOCUS 166937 bp DNA linear HTG 06-MAY-2001
DEFINITION Homo sapiens chromosome 15 clone RP11-109D20 map 15, WORKING DRAFT
SEQUENCE, 22 unordered pieces.
AC012255
ACCESSION AC012255
VERSION AC012255.4 GI:7684440
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 166937)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-109D20
Unpublished
2 (bases 1 to 166937)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckert,Y.R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heathford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., MacDonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Polara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission

TITLE
JOURNAL
COMMENT
Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 3, 2000 this sequence version replaced gi:7209928.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L3737

Center clone name: 109 D_20

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 153800 bases at least Q40

Consensus quality: 160606 bases at least Q30

Consensus quality: 163275 bases at least Q20

Insert size: 164837; sum-of-contigs

Quality coverage: 4.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 2087: contig of 2087 bp in length
* 2088 2187: gap of 100 bp
* 2188 3626: contig of 1439 bp in length
* 3627 3726: gap of 100 bp
* 3727 4845: contig of 1119 bp in length
* 4846 4945: gap of 100 bp
* 4946 6702: contig of 1757 bp in length
* 6703 6802: gap of 100 bp
* 6803 8326: contig of 1524 bp in length
* 8327 8426: gap of 100 bp

* 8427 10494: contig of 2068 bp in length
* 10495 10594: gap of 100 bp
* 10595 13215: contig of 2621 bp in length
* 13216 13315: gap of 100 bp
* 13316 16451: contig of 3136 bp in length
* 16452 16551: gap of 100 bp
* 16552 20843: contig of 4292 bp in length
* 20844 20943: gap of 100 bp
* 20944 24940: contig of 3997 bp in length
* 24941 25040: gap of 100 bp
* 25041 29901: contig of 4861 bp in length
* 29902 30001: gap of 100 bp
* 30002 34189: contig of 4188 bp in length
* 34290 34289: gap of 100 bp
* 34290 38386: contig of 4097 bp in length
* 38387 38486: gap of 100 bp
* 38487 45827: contig of 7341 bp in length
* 45828 45927: gap of 100 bp
* 45928 51729: contig of 5802 bp in length
* 51730 51829: gap of 100 bp
* 51830 60123: contig of 8294 bp in length
* 60124 60223: gap of 100 bp
* 60224 73532: contig of 13309 bp in length
* 73533 73632: gap of 100 bp
* 73633 87433: contig of 13801 bp in length
* 87434 87533: gap of 100 bp
* 87534 101708: contig of 14175 bp in length
* 101709 101808: gap of 100 bp
* 101809 121193: contig of 19385 bp in length
* 121194 121293: gap of 100 bp
* 121294 139051: contig of 17758 bp in length
* 139052 139151: gap of 100 bp
* 139152 166937: contig of 27786 bp in length.

FEATURES
source
Location/Qualifiers

1. 166937
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-109D20"
/clone_lib="RP11-11 Human Male BAC"
1. 2087
/note="assembly_fragment"
2188. 3626
/note="assembly_fragment"
3727. 4845
/note="assembly_fragment"
4946. 6702
/note="assembly_fragment"
6803. 8326
/note="assembly_fragment"
8427. 10494
/note="assembly_fragment"
10595. 13215
/note="assembly_fragment"
13316. 16451
/note="assembly_fragment"
16552. 20843
/note="assembly_fragment"
20944. 24940
/note="assembly_fragment"
25041. 29901
/note="assembly_fragment"
30002. 34189
/note="assembly_fragment"
34290. 38386
/note="assembly_fragment"
38487. 45827
/note="assembly_fragment"
45928. 51729
/note="assembly_fragment"
51830. 60123

finished clone CTD-2651B20, AC051619 [Drafting center: UWMSC] in order to give overlap required for determining long range contiguity.

FEATURES

source

Location/Qualifiers

1..181312

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="15"

/map="15q15"

/clone="RP11-109D20"

/clone_lib="RPCI human BAC library 11"

/note="This clone overlaps CTD-2014N11 AC090888 and

RP11-276K9 AC087790. Data from overlapping BACs were added

and the consensus sequence determined from RP11-109D20 to

the extent possible."

1..89137

/note="overlap with CTD-2014N11, AC090888"

36107..36120

/note="low quality data"

40470..40820

/note="sequence data generated from subcloned PCR product"

53030..53480

/note="sequence data generated from subcloned PCR product"

86982..86983

/note="low quality data"

95610..96030

/note="sequence data generated from subcloned PCR product"

96830..97420

/note="sequence data generated from subcloned PCR product"

98260..98420

/note="single clone coverage"

98850..98852

/note="low quality data"

100981..101003

/note="low quality data"

174237..179372

/note="overlap with RP11-276K9, AC087790. This sequence

data was added to establish the overlap between the BACs

in the tiling path"

179373..181312

/note="overlap with CTD-2651B20, AC051619. This data is

added to give overlap for long range contiguity."

misc_feature

misc_feature

unsure

unsure

misc_feature

misc_feature

misc_feature

unsure

misc_feature

misc_feature

unsure

misc_feature

ORIGIN

Query Match

Best Local Similarity 78.7%; Score 392; DB 9; Length 181312;

Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

91 GAGCTCCTGCTCTTCTCTCGAGAGGCGCTGTGAGTCTCCAGATGTTGGGCCAGC 150

Db

168383 GAGCTCCTGCTCTTCTCTCGAGAGGCGCTGTGAGTCTCCAGATGTTGGGCCAGC 168442

QY

151 GCTCTTGCACCTTCTTGACCAAGCGCAAGAGTGCAGCCAGAGAGAGGCGCTCA 210

Db

168443 GCTCTTGCACCTTCTTGACCAAGCGCAAGAGTGCAGCCAGAGAGAGGCGCTCA 168502

QY

211 CCTCTTATCTCGGCGACCACTGCACAGCAGCGCTCTCCAGACTTAAATGTATC 270

Db

168503 CCTCTTATCTCGGCGACCACTGCACAGCAGCGCTCTCCAGACTTAAATGTATC 168562

QY

271 ACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCCTTGGAGCATGCGAG 330

Db

168563 ACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCCTTGGAGCATGCGAG 168622

QY

331 CCGGGAAGCAGTGCCTCGGCGAGGCTGGGCGAGAGAGCTCCAGGAAAGGCACTGAGCGCT 390

Db

168623 CCGGGAAGCAGTGCCTCGGCGAGGCTGGGCGAGAGAGCTCCAGGAAAGGCACTGAGCGCT 168682

QY

391 GCTGGCGGAGGCTCTGGACATCCGAGGAGCAAGGAAAGTCTCTGGGGCGATCTGTA 450

Db

168683 GCTGGCGGAGGCTCTGGACATCCGAGGAGCAAGGAAAGTCTCTGGGGCGATCTGTA 168742

QY 451 AATAAACCTTTTCTTTGTTTAAAAA 482
Db 168743 AATAAACCTTTTCTTTGTTTAAAAA 168774

RESULT 9

HSM806241

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

ORIGIN

polyA_signal

polyA_site

Query Match

Best Local Similarity 75.1%; Score 374; DB 9; Length 1740;

Matches 424; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

62 GCGCGCCTTGTGGTACCGCTGCAACCGCGCTGCTGCTCTTCTCTCGAGGGCGG 121

Db

1305 GCGCGCCTTGTGGTACCGCTGCAACCGCGCTGCTGCTCTTCTCTCGAGGGCGG 1364

QY

122 TGGTGAAGTCTCAGTATGTTGCGCCAGCGCTCTTGCACCTTCTGGAACCAAGCGCCA 181

Db

1365 TGGTGAAGTCTCAGTATGTTGCGCCAGCGCTCTTGCACCTTCTGGAACCAAGCGCCA 1424

QY

182 AGGACTGCAGGAGAGAGAGAGGGGCTACCTTATCTCTGCGAGCCACTGCACAAGC 241

Db

1425 AGGACTGCAGGAGAGAGAGAGGGGCTACCTTATCTCTGCGAGCCACTGCACAAGC 1484

QY

242 AGGCGGCTCTCCAGACTTAAATGTATACCACTAACCCTGTGAGGGGAGCCCAATCTGG 301

Db

1485 AGGCGGCTCTCCAGACTTAAATGTATACCACTAACCCTGTGAGGGGAGCCCAATCTGG 1544

QY

302 ACTCCTTCCCGCCTTGGAGCATCGCAGGCGGGAAGCAGTGGCCCGCAGGCGCTGGGCA 361

Db

1545 ACTCCTTCCCGCCTTGGAGCATCGCAGGCGGGAAGCAGTGGCCCGCAGGCGCTGGGCA 1604

QY

362 GGAAGCTCCAGGAAGGCACTGAGCGCTGCTGGCGGAGGCTCGGACATCCGAGGCA 421

Db 1605 GGAGAGCTCCAGGAAGGCACTGAGCGCTGTCGGCGGAGGCTTCGACATCCGACAGCA 1664

QY 422 CCAGGAAAGTCTCTCGGGCGATCTGTAATAAACCCTTTTCTTTGTTTAAAA 481

Db 1665 CCAGGAAAGTCTCTCGGGCGATCTGTAATAAACCCTTTTCTTTGTTTAAAA 1724

QY 482 AAAA 486

Db 1725 AAAA 1729

RESULT 10

HS183H12F/c 217 bp DNA linear PRI 18-OCT-1995

LOCUS H.sapiens Cpg island DNA genomic MseI fragment, clone 183h12,

DEFINITION forward read cpg183h12.ftla.

ACCESSION 257552

VERSION 257552.1 GI:1028783

KEYWORDS Cpg island; genomic MseI fragment.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.

TITLE Purification of Cpg islands using a methylated DNA binding column

JOURNAL Nat. Genet. 6 (3), 236-244 (1994)

MEDLINE 94282070

PUBMED 8012384

REFERENCE 2 (bases 1 to 217)

AUTHORS Dodsworth, S.J., Huckle, E., Wilkinson, P. and Micklem, G.

TITLE Direct Submission

JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire, CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk

COMMENT Vector: pGEM-5Zf(-)

Clones are available from the UK MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL: <http://www.hgmp.mrc.ac.uk/> for details or contact: biohelp@hgmp.mrc.ac.uk.

FEATURES

source

1. .217

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/clone="183h12"

/sex="male"

/tissue_type="blood"

/clone_id="CGI-1"

/dev_stage="adult"

ORIGIN

Query Match 43.6%; Score 217; DB 9; Length 217;

Best Local Similarity 100.0%; Pred. No. 1.6e-110;

Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 AAATGTATCACCACTAATCTGAGGGGAGCCCAATCTGACTCTCCCGCCTTGGGA 321

Db 217 AAATGTATCACCACTAATCTGAGGGGAGCCCAATCTGACTCTCCCGCCTTGGGA 158

QY 322 CATCGAGCCGGGAAGCAGTGCCTCCGACCTGGGCCAGAGAGCTCCAGGAAGGCA 381

Db 157 CATCGAGCCGGGAAGCAGTGCCTCCGACCTGGGCCAGAGAGCTCCAGGAAGGCA 98

QY 382 CTGAGCGCTGCTGGCGCGAGGCTTCGACATCCGAGGACCCAGGAAAGTCTCTCTGGGG 441

Db 97 CTGAGCGCTGCTGGCGCGAGGCTTCGACATCCGAGGACCCAGGAAAGTCTCTCTGGGG 38

QY 442 CGATCTGTAATAAACCTTTTCTTTGTTTGA 478

Db 37 CGATCTGTAATAAACCTTTTCTTTGTTTGA 1

RESULT 11

BC029819/c

LOCUS BC029819 1923 bp mRNA linear PRI 06-OCT-2003

DEFINITION Homo sapiens hypothetical protein FLJ32334, mRNA (cDNA clone MGC:35310 IMAGE:5177619), complete cds.

ACCESSION BC029819

VERSION BC029819.1 GI:20987583

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1923)

AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalhus, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

MEDLINE 12477932

PUBMED 12477932

REFERENCE 2 (bases 1 to 1923)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK

COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabs-rc@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louisged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Navavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAC Plate: 51 Row: 0 Column: 3

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21389332.

FEATURES

source

1. .1923

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:35310 IMAGE:5177619"

/tissue_type="Brain, Lung, Testis, adult, pooled whole"

/clone_id="NIH_MGC_115"

/lab_host="DH10B"

/note="Vector: PCMV-SPORT6"

1. .1923

gene

/gene="FLJ32334"
/db_xref="LocusID:90527"
401.1852
/codon_start=1
/product="FLJ32334 protein"
/protein_id="AAH29819.1"
/db_xref="GI:20987584"
/db_xref="LocusID:90527"
/translation="MAYLGHFPFYAGKPTFPMDTTASIMIFLTALATFVILPG
IRGKTRLFMLRVVTSIFGAALAVNFSSEMSVGVSTNTSYKAFSSSEWISADIGLO
VGLGVNITLTPTVQQLNETINNEFTWRLGENYAEYAKALEKGLPDPVLYLAEK
FTPRSPCGYRQYRLAGHTSAMLVAFVLCMLANVMSPLVYGGYMLATGIFOL
LALLFSPMATSLSPCPLHIGASVLTHTHGPAPFWITLTGLCYLIGLMAVAARMQ
HRLKAFNOSVDEDPMLESPPEEGILSPRYSMADSPKSDIPLEASSTKAYRPR
RLSLVPADYRGLAPALSLPGALLAQWRALLPGLRCPKAKESRLGPHSPWRFGP
EGCEERWAEHTGDSPRPLRGRTGRLWRWGSKERACGVRAMLPRLVSNGLKRPSC
LDPKCMYRRDARAFPHLEPTPCVTSRHTPLI"

ORIGIN

Query Match 33.3%; Score 166; DB 9; Length 1923;
Best Local Similarity 100.0%; Pred. No. 1.1e-81;
Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 CTGTGAGGGGGAACCAATCTGACTCCTTCCCGCTTGGGACATCGACGGCCGGGAAGC 339
DB 1557 CTGTGAGGGGGAACCAATCTGACTCCTTCCCGCTTGGGACATCGACGGCCGGGAAGC 1498
QY 340 AGTCCCCCGGAGCCCTGGGCGCAGAGAGCTCCAGGAAGGCACTGACCGCTGCGCGCG 399
DB 1497 AGTCCCCCGGAGCCCTGGGCGCAGAGAGCTCCAGGAAGGCACTGACCGCTGCGCGCG 1438
QY 400 AGGCCTCGGACATCCGCGAGCACCAGGAAAGTCTCTGGGGCGCAT 445
DB 1437 AGGCCTCGGACATCCGCGAGCACCAGGAAAGTCTCTGGGGCGCAT 1392

RESULT 12
HS183A12R 218 bp DNA linear PRI 19-OCT-1995
LOCUS H.sapiens Cpg island DNA genomic MseI fragment, clone 183a12,
DEFINITION reverse read cpg183a12.r1a.
ACCESSION Z59954
VERSION Z59954.1 GI:1031867
KEYWORDS Cpg island; genomic MseI fragment.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
TITLE Purification of Cpg islands using a methylated DNA binding column
JOURNAL Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE 94282070
PUBMED 8012384
REFERENCE 2 (bases 1 to 218)
AUTHORS Dodsworth,S.J., Huckle,E., Wilkinson,P. and Micklem,G.
TITLE Direct Submision
JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk
COMMENT Clones are available from the UK MRC Human Genome Mapping Project
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
http://www.hgmp.mrc.ac.uk/ for details
or contact: biochelp@hgmp.mrc.ac.uk.
FEATURES
Source 1.218
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="183a12"
/sex="male"
/tissue type="blood"
/clone_1ib="CGI-1"

/dev_stage="adult"

ORIGIN
Query Match 27.3%; Score 136; DB 9; Length 218;
Best Local Similarity 100.0%; Pred. No. 7.6e-65;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 CGGAGACAGTGCCTCCGAGGCTGGGCGCAGAGAGCTCCAGAGAGGCACTGAGCGCTG 391
DB 72 CGGAGACAGTGCCTCCGAGGCTGGGCGCAGAGAGCTCCAGAGAGGCACTGAGCGCTG 131
QY 392 CTGGCGGAGGCTCGGACATCCGCGAGGACCCAGGAAAGTCTCTGGGGCGATGTAA 451
DB 132 CTGGCGGAGGCTCGGACATCCGCGAGGACCCAGGAAAGTCTCTGGGGCGATGTAA 191
QY 452 ATAAACCTTTTCT 467
DB 192 ATAAACCTTTTCT 207

RESULT 13
AC069544 214866 bp DNA linear PRI 25-APR-2002
LOCUS AC069544
DEFINITION Homo sapiens chromosome 10 clone RP11-398C13, complete sequence.
ACCESSION AC069544
VERSION AC069544.9 GI:20303530
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 214866)
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 214866)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 3 (bases 1 to 214866)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2001) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 4 (bases 1 to 214866)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On Apr 25, 2002 this sequence version replaced gi:117975233.
FEATURES
Source 1.214866
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-398C13"
/clone_1ib="RPCT-11"

ORIGIN
Query Match 5.2%; Score 26; DB 9; Length 214866;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 464 TTCTTTGTTTAAAAAAATTA 489
DB 200470 TTCTTTGTTTAAAAAAATTA 200495
RESULT 14
AE014827

LOCUS AE014827 254436 bp DNA linear INV 07-OCT-2002
DEFINITION Plasmodium falciparum 3D7 chromosome 14 section 12 of 13 of the
complete sequence.
ACCESSION AE014827 AE014187
VERSION AE014827.1 GI:23497719
KEYWORDS
SOURCE Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
REFERENCE 1 (bases 1 to 254436)
Fukuyota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS Gardner,M.J., Hall,N., Fung,E., White,O., Beriman,M., Hyman,R.W.,
Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
Kyes,S., Chan,M.-S., Nene,V., Shalom,S.J., Suh,B., Peterson,J.,
Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
Barrell,B.
TITLE Genome sequence of the human malaria parasite Plasmodium falciparum
JOURNAL Nature 419 (6906), 498-511 (2002)
PUBMED 12368864
REFERENCE 2 (bases 1 to 254436)
AUTHORS Gardner,M.J.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source 1..254436
/organism="Plasmodium falciparum 3D7"
/mol_type="genomic DNA"
/isolate="3D7"
/db_xref="taxon:36329"
/chromosome="14"
repeat_region complement(68..243)
/rpt_type=tandem
/rpt_unit="(TA)n"
68..143
/rpt_type=tandem
68..132
repeat_region /rpt_type=tandem
200..243
repeat_region /rpt_type=tandem
complement(295..385)
/note="AT-rich"
repeat_region /rpt_type=tandem
complement(431..521)
/rpt_type=tandem
/rpt_unit="(TAAAAA)n"
complement(527..555)
/note="AT-rich"
repeat_region /rpt_type=tandem
complement(582..607)
/note="AT-rich"
repeat_region /rpt_type=tandem
608..710
/rpt_type=tandem
/rpt_unit="(TAA)n"
complement(794..822)
/note="AT-rich"
repeat_region /rpt_type=tandem
complement(839..889)
/note="AT-rich"
repeat_region /rpt_type=tandem
complement(946..995)
/note="AT-rich"
repeat_region /rpt_type=tandem
complement(1002..1034)
/rpt_type=tandem
/rpt_unit="(TA)n"
1002..1031
repeat_region /rpt_type=tandem

repeat_region complement(1057..1106)
/rpt_type=tandem
/rpt_unit="(A)n"
repeat_region complement(1112..1146)
/note="AT-rich"
/rpt_type=tandem
complement(1204..1362)
/note="AT-rich"
repeat_region /rpt_type=tandem
1382..1405
/rpt_type=tandem
/rpt_unit="(A)n"
complement(1406..1428)
/note="AT-rich"
repeat_region /rpt_type=tandem
complement(1447..1544)
/note="AT-rich"
repeat_region /rpt_type=tandem
complement(1545..1568)
/rpt_type=tandem
/rpt_unit="(TA)n"
complement(1644..1690)
/note="AT-rich"
repeat_region /rpt_type=tandem
1703..2211
/locus_tag="PF14_0645"
join(<1703..1751,1852..1964,2125..>2211)
/locus_tag="PF14_0645"
join(1703..1751,1852..1964,2125..2211)
/locus_tag="PF14_0645"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA037258.1"
/db_xref="GI:23497720"
/translation="MENIETMKPYFPASINGCESVSDEFKCLNKNLIPFGDDKLKS
SQDCCYFKNYEKCYDEKLKLTPLMFLTEYKKNK"
complement(1764..1846)
/note="AT-rich"
repeat_region /rpt_type=tandem
2044..2092
/rpt_type=tandem
complement(2048..2092)
/rpt_type=tandem
/rpt_unit="(TA)n"
complement(2097..2123)
/note="AT-rich"
repeat_region /rpt_type=tandem
complement(2239..2366)
/note="AT-rich"
repeat_region /rpt_type=tandem
complement(2367..2396)
/rpt_type=tandem
/rpt_unit="(A)n"
complement(2398..2418)
/rpt_type=tandem
/rpt_unit="(TA)n"
complement(2469..2764)
/note="AT-rich"
repeat_region /rpt_type=tandem
2822..2865
/rpt_type=tandem
complement(2822..2865)
/rpt_type=tandem
/rpt_unit="(TA)n"
complement(2904..2927)
/note="AT-rich"
repeat_region /rpt_type=tandem
complement(2950..2974)
/note="AT-rich"
repeat_region /rpt_type=tandem
complement(3003..3059)
/note="AT-rich"
repeat_region /rpt_type=tandem

```
repeat_region      complement(3093. .3152)
                    /note="AT_rich"
                    /rpt_type=tandem
gene                <3157. .>3530
                    /locus_tag="PF14_0646"
                    join(<3157. .3281,3389. .>3530)
mRNA                /locus_tag="PF14_0646"
                    join(3157. .3281,3389. .3530)
CDS                 /locus_tag="PF14_0646"
                    /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AAN37259.1"
                    /db_xref="GI:23497721"
                    /translation="MFTARKDFNDYKICMQSHLNKDIACEKCELYKAINSTSHIIS
RECLPYTEDLOKCFKHSFRLSFCDKEIMDKLKTCCQSDVYNLITS"
repeat_region      3315. .3366
                    /rpt_type=tandem
                    complement(3315. .3366)
repeat_region      /rpt_type=tandem
                    /rpt_unit="(TA)n"
repeat_region      complement(3533. .3596)
                    /note="AT_rich"
                    /rpt_type=tandem
                    complement(3597. .3638)
repeat_region      /rpt_type=tandem
                    /rpt_unit="(A)n"
                    /rpt_unit="(A)n"
                    3613. .3638
repeat_region      /rpt_type=tandem
                    complement(3713. .3789)
                    /note="AT_rich"
                    /rpt_type=tandem
                    complement(3859. .3895)
repeat_region      /note="AT_rich"
                    /rpt_type=tandem
                    3926. .3961
repeat_region      /rpt_type=tandem
                    complement(3926. .3957)
                    /rpt_type=tandem
                    /rpt_unit="(TA)n"
                    complement(3997. .4137)
repeat_region      /rpt_type=tandem
                    /rpt_unit="(TA)n"
                    4194. .4227
repeat_region      /rpt_type=tandem
                    /rpt_unit="(A)n"
                    complement(4286. .4320)
repeat_region      /rpt_type=tandem
                    /rpt_unit="(TAAAA)n"
                    complement(4335. .4398)
                    /note="AT_rich"
                    /rpt_type=tandem
                    4529. .4555
repeat_region      /rpt_type=tandem
                    /rpt_unit="(A)n"

Query Match      5.2%; Score 26; DB 3; Length 254436;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY                466 CTTTGTGTTTAAAAAAATAAAA 491
Db                60242 CTTTGTGTTTAAAAAAATAAAA 60267

RESULT 15
LOCUS      AB025602                55790 bp      DNA      linear      PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, BAC clone:F14A1.
ACCESSION  AB025602 BA000015
VERSION     AB025602.1 GI:4589408
KEYWORDS
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM    Arabidopsis thaliana
```

```
REFERENCE
AUTHORS      Kaneko,T., Katoh,T., Asamizu,E., Sato,S., Nakamura,Y., Kotani,H.
              and Tabata,S.
TITLE        Structural analysis of Arabidopsis thaliana chromosome 5. XI
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 55790)
AUTHORS      Nakamura,Y.
TITLE        Direct Submision
JOURNAL      Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
              Institute, Department of Plant Gene Research, 1532-3, Yana,
              Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp,
              Tel:81-438-52-3935, Fax:81-438-52-3934)
              Address for correspondence: kaos@kazusa.or.jp
              For the latest information on annotation of this clone, please see
              http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=F14A1
              Genes with similarity to proteins in the databases are described in
              'product' or 'note' qualifiers. Genes that have no significant
              protein similarity are described as 'unknown protein'.
              The software programs used to predict genes include: Grail
              (Informatics Group, Oak Ridge National Laboratory,
              http://compbio.ornl.gov/Grail-1.3/),
              GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
              NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
              Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
              SplicePredictor (Volker Brendel, Stanford University,
              http://gremliml.2001.fasstate.edu/cgi-bin/sp.cgi).
              Genes encoding tRNAs are predicted by tRNAscan-SE
              (Sean Eddy, Washington University School of Medicine, St. Louis,
              http://genome.wustl.edu/eddy/tRNAscan-SE/).
              This sequence may not be the entire insert of this clone. It may be
              shorter because we remove overlaps between neighboring submissions.
              The 5' clone is M1K22 and the 3' clone is MBE13.

FEATURES
source        1. 55790
               /organism="Arabidopsis thaliana"
               /mol_type="genomic DNA"
               /strain="Columbia"
               /db_xref="taxon:3702"
               /chromosome="5"
               /clone="F14A1"
               /clone_1ib="IGF BAC"
               complement(345. .538)
               /product="nucleolar protein-like"
               /note="CDS is reported in Acc# AB005236
               gene_id:M1K22.22"
               /number=2
               /evidence=not experimental
               complement(650. .1037)
               /product="nucleolar protein-like"
               /note="CDS is reported in Acc# AB005236
               gene_id:M1K22.22"
               /number=1
               /evidence=not experimental
               complement(2055. .2228)
               /note="gb|AAD23664.1
               gene_id:F14A1.2
               similar to unknown protein"
               /pseudo
               /codon_start=1
               /evidence=not experimental
               complement(join(3022. .3227,3496. .4530))
               /note="contains similarity to unknown protein
               gb|AAD38227.1
               gene_id:F14A1.3"
               /pseudo
               /codon_start=1
               /evidence=not experimental
               join(8265. .8758,8827. .9079,10490. .10842,10919. .11470)
               /note="gene_id:F14A1.4
               pir||T01797
```

strong similarity to unknown protein"
/pseudo
/codon_start=1
/evidence=not_experimental
join(14755..15729,16136..16766)
/note="contains similarity to cytochrome P450
gene_id:F14A1.5"
/pseudo
/codon_start=1
/evidence=not_experimental
17455..18278
/note="contains similarity to Athila ORF 1
gene_id:F14A1.6"
/pseudo
/codon_start=1
/evidence=not_experimental
22069..22536
/note="gene_id:F14A1.7"
/codon_start=1
/evidence=not_experimental
/product="cytochrome P450-like"
/protein_id="BAA97454.1"
/db_xref="GI:8885524"
/translation="MKKLKILKEAFRLHHVAPENLPHMSTTDVVDGYFIPKGSKLI
SRMGIGRNPKVMDKPLKFDPERHLSNNTCVLKEPDNITTFGAERRGCMGADIGSAM
TYMLVYARLIQGFMTSPVSESKIDISVSKSDLFMAKPLNAVVTPLDENVYPA"
24745..25395
/note="contains similarity to Athila retroelement ORF1
protein
gene_id:F14A1.8"
/pseudo
/codon_start=1
/evidence=not_experimental
join(29275..29414,29688..30532,30645..30785,30870..31208)
/note="gb|AAD29066.1
gene_id:F14A1.9
similar to unknown protein"
/pseudo
/codon_start=1
/evidence=not_experimental
complement(join(32433..32517,32622..32684,32833..33045,
33139..33248,33712..33756,33846..33905,34042..34091,
34183..34267,34610..34669,34811..34919,35405..35448,
35606..35692,35775..35850,36080..36744,37204..37923,
38243..38449,38897..39093,39661..40059,40141..40393))
/note="gene_id:F14A1.10
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA97455.1"
/db_xref="GI:8885525"
/translation="MSERNRGKPRNCCISHKPYDAASKNPHQIAVIOAGSLTARSIRS
ESAAPVYDGDCKPTFAELSSIDSLTSLRLHRLSNDSDVAGSDVVMKVALYMP
PSVEYIVSVESVLRGGEAFPLDPSWPRERVLSSISNLSIVIACGLSVESHWLVE
RNVCPLYLFSMDEKLSVEFGSSFWPCCKEQRKFCYLMYTSSTGKPKGVCGETEOQ
LNRFLMQLYPVVGEQRFPAFTSVGFIDHTIQEFLGALISSTALVIPPFTLLKENMI
SIIDFEEYISIRLAVPSMIRALPTLQHRGNKKLQSCUKLVLSGEPFVSLWDS
LHSLPETCFINLYGSTSVSGDCCTYFDCSELPRLLKTEEIGSVPIGKSISNCKVLLG
DEDKPYEGEICVSGCLISQGYMHSISESGYKLNHNSLCNHLTNDGSDLIYRTGDY
GROLSSGDLIFIGRDRTVKLNGRMALBEIETLELNPDIABAVLLSRDETELASL
KAFVTLNKSNSSDGIIFSIRNMWGKLPVMI PNHFVLEKLP LTTSSGKYDYEALAR
LKCPITGAODMMOSNGTNSLNONIKKAVCDALVKEVSDDDFEAIIGDSLAAHLISH
SLGIDMRLLIYQFRSPSRLLIYLSEKEGLREDMOHNTQKLDHKIESQNGLVSRIV
PLHSGVTSPTPSKLOCEKNNSPKRLKIDYEKFSPKRMKENKLWDSGSOIQCAFSRC
NKVHSPESCNEANREYWSLEI PRNQVMQETMKVHMESCVDA SPLVLKDSKTYL
FIGSHRKFSCIDAKSGSMYETILLEGSAWVDFSVVIGCYGKALYFLDFST
GSLCMKFOACGEIKCPVVDITSSQLIWGSHDHTLYALDPRQCCVYKLOCGGSIFAS
PAIDEGHSLYVASTSGRVIASIKDSPFHTLMLELEAPIFGSLCITPTQNVDCV
IAMSPTGTLIWRRTGGRPIFAGPCMSHVLPSQYLVCCRNGCVYSLPEBSGLVWEDNI
GDPITASAYIDENLHFSHELASDLVTVCSSGRVHVLVRPSSILSRSHDSKVGE
ITRMEIADIFSSPVMIGRIFVGCDDYVHCLSLBESCRTTTFEKRIMTSLDIKFF
LELYELETGTKIVIIKEKHNGTNNLI"
join(41918..41972,42875..42975,43101..43682)

CDS

CDS

CDS

CDS

/note="gb|AAD15520.1
gene_id:F14A1.11
similar to unknown protein"
/pseudo
/codon_start=1
/evidence=not_experimental
complement(join(49029..49804,49883..50017,50099..51775,
51848..52168,52220..52875,52933..53719))
/note="contains similarity to retroelement pol polyprotein
gene_id:F14A1.12"
/pseudo
/codon_start=1
/evidence=not_experimental
complement(54463..55327)
/note="gene_id:F14A1.13
pir||T05807
similar to unknown protein"
/pseudo
/codon_start=1
/evidence=not_experimental

CDS

ORIGIN

Query Match	5.0%;	Score 25;	DB 8;	Length 55790;
Best Local Similarity	100.0%;	Pred. No. 0.048;		
Matches	25;	Conservative 0;	Mismatches 0;	Indels 0;
				Gaps 0;

QY 459 TTTTTCCTTTGTTTTTTAAAAA 483
|||||
Db 20789 TTTTTCCTTTGTTTTTTAAAAA 20765

Search completed: February 24, 2004, 01:57:08
Job time : 1969.03 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 19:50:14 ; Search time 201.357 Seconds
(without alignments)
10506.704 Million cell updates/sec

Title: US-09-936-456-3

Perfect score: 498

Sequence: 1 ctctagcgtgcgcgtctgcc.....aaaaataataaagtcgacc 498

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

1: geneseqn1980s:*

2: geneseqn1990s:*

3: geneseqn2000s:*

4: geneseqn2001as:*

5: geneseqn2001bs:*

6: geneseqn2002s:*

7: geneseqn2003as:*

8: geneseqn2003bs:*

9: geneseqn2003cs:*

10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	498	100.0	498	3	AAA94624	Aa94624 Human CAS
2	425	85.3	1441	3	AAA94623	Aa94623 Human CAS
3	425	85.3	1474	6	AB190535	AB190535 Human pol
4	425	85.3	1491	9	ADD19228	Add19228 Human CDN
5	416	83.5	1421	3	AAA96505	Aa96505 CDNA enco
6	392	78.7	2684	7	ADA53627	Ada53627 Human cod
7	392	78.7	5033	7	AB268115	Ab268115 Human sec
8	392	78.7	5033	7	AB274587	Ab274587 Secreted
9	392	78.7	5033	9	ADC21005	Adc21005 Human sec
10	392	78.7	5033	9	ADC21005	Adc21005 Human sec
11	228	45.8	1460	7	AB267235	Ab267235 Human sec
12	228	45.8	1460	7	AB273640	Ab273640 Secreted
13	228	45.8	1460	9	ADC20289	Adc20289 Human sec
14	26	5.2	525	4	AAK81950	Aak81950 Human imm
15	26	5.2	525	4	AAK81951	Aak81951 Human imm
16	26	5.2	525	4	AAK81952	Aak81952 Human imm
17	23	4.6	535	4	AAK62215	Aak62215 Human imm
18	22	4.4	900	5	AAS68034	Aas68034 DNA enco
19	22	4.4	3243	4	ABK42289	Abk42289 Genomic s
20	22	4.4	3243	8	ADB60445	Adb60445 Connectiv
21	21	4.2	6415	2	AAS46371	Aas46371 Tumour su
22	21	4.2	299	2	AAT19557	Aat19557 Human gen
23	21	4.2	325	7	ABX45555	Abx45555 Bovine ES

C	24	21	4.2	335	5	ABV47656	Abv47656 Human pro
	25	21	4.2	569	4	AA187726	Aa187726 Human pol
	26	21	4.2	569	8	ACH15807	Ach15807 Human adu
C	27	21	4.2	623	3	AAA16202	Aa16202 Human col
	28	21	4.2	1418	6	ABS65211	Ab65211 CDNA enco
	29	21	4.2	1420	7	ABV72442	Abv72442 Nucleotid
	30	21	4.2	1473	9	ADE25603	Ade25603 Human CDN
	31	21	4.2	1484	3	AAF15813	Aaf15813 Human pro
	32	21	4.2	1505	2	AAZ41298	Aaz41298 Human nor
	33	21	4.2	1558	6	AB665212	Ab665212 CDNA enco
C	34	21	4.2	2573	4	AAK82049	Aak82049 Human imm
	35	21	4.2	3730	6	ABV95105	Abv95105 Gene #160
	36	21	4.2	13084	6	AB134179	Ab134179 Human imm
	37	21	4.2	17703	6	ABK39952	Abk39952 Human che
	38	21	4.2	18060	6	AB192213	Ab192213 Chemicall
	39	21	4.2	18060	6	AAD22314	Aad22314 Chemicall
	40	21	4.2	110000	6	ABA90521_03	Continuation (4 of
	41	20	4.0	21	4	AAF96883	Aaf96883 Human gen
	42	20	4.0	51	4	AA131676	Aa131676 Human SNP
	43	20	4.0	282	5	ABV61253	Abv61253 Human pro
	44	20	4.0	303	5	ABV61433	Abv61433 Human pro
C	45	20	4.0	308	5	ABV60729	Abv60729 Human pro

ALIGNMENTS

RESULT 1

AAA94624

ID AAA94624 standard; DNA; 498 BP.

XX

AC AAA94624;

XX

DT 11-JAN-2001 (first entry)

XX

DE Human CASB618 EST.

XX

KW Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian;

XX

KW colon; autoimmune disease; HLA_A0201; expressed sequence tag; EST; ss.

XX

OS Homo sapiens.

XX

PN WO200053748-A2.

XX

PD 14-SEP-2000.

XX

PF 09-MAR-2000; 2000WO-EP002048.

XX

PR 11-MAR-1999; 99GB-00005607.

PR 01-SEP-1999; 99GB-00020590.

XX

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX

PI Bruck CEM, Cascart J, Coche T, Vinals Y De BassolsC;

XX

DR WPI; 2000-572268/53.

XX

PT New human CASB618 polypeptide, useful as a vaccine for prophylactic and

PT therapeutic treatment of cancers, particularly ovarian or colon cancer,

PT autoimmune diseases and related conditions.

XX

PS Claim 32; Page 62; 76pp; English.

XX

CC The present sequence is an expressed sequence tag (EST) for human CASB618

CC protein. The gene for human CASB618 is thought to be located on

CC chromosome 15. CASB618 protein and epitopes of CASB618 protein (see

CC AAB26327 to AAB26399) are useful in diagnosing the occurrence of tumour

CC cells and in vaccines for prophylactic and therapeutic treatment of

CC cancers, particularly ovarian or colon cancer, autoimmune diseases and

CC related conditions

XX

SQ Sequence 498 BP; 97 A; 166 C; 137 G; 98 T; 0 U; 0 Other;

```
Query Match      100.0%; Score 498; DB 3; Length 498;
Best Local Similarity 100.0%; Pred. No. 3.8e-238;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	1	CTCTAGCGTGCCCGCTCTGCCCCGCTCCCGCTTAGAGCTCCTCCGGGCTCAACCACTCAGTACG	60
Db	1	CTCTAGCGTGCCCGCTCTGCCCCGCTCCCGCTTAGAGCTCCTCCGGGCTCAACCACTCAGTACG	60
QY	61	AGCGCCGCTTCTGGGTCAACGCTGGCAACCGGCGTCTGTGCTCTTCTCTGGAGGGGCC	120
Db	61	AGCGCCGCTTCTGGGTCAACGCTGGCAACCGGCGTCTGTGCTCTTCTCTGGAGGGGCC	120
QY	121	GTGTTGATCTCTCCAGTATGTTCCGGCCACGCGCTCTTCCGACCCCTTCTGACCAAGCGCC	180
Db	121	GTGTTGATCTCTCCAGTATGTTCCGGCCACGCGCTCTTCCGACCCCTTCTGACCAAGCGCC	180
QY	181	AAGACTGCAGCCAGGAGAGAGGGGGCTCACCCTTATCTCGGGGACCCCACTGCACAAG	240
Db	181	AAGACTGCAGCCAGGAGAGAGGGGGCTCACCCTTATCTCGGGGACCCCACTGCACAAG	240
QY	241	CAGCGCGCTCTCCAGACTTAAATGTATCACCACTAACCTGTGAGGGGACCCAATCTG	300
Db	241	CAGCGCGCTCTCCAGACTTAAATGTATCACCACTAACCTGTGAGGGGACCCAATCTG	300
QY	301	GACTCCTTCCCGCCTTGGGACATCGCAGGCGGGGAAAGCAGTGCCTCGCAGGCTGAGGC	360
Db	301	GACTCCTTCCCGCCTTGGGACATCGCAGGCGGGGAAAGCAGTGCCTCGCAGGCTGAGGC	360
QY	361	AGGAGAGCTCCAGGAAAGGACACTGACGCTGTGGCGCGAGGCTTCGACATCCGAGGC	420
Db	361	AGGAGAGCTCCAGGAAAGGACACTGACGCTGTGGCGCGAGGCTTCGACATCCGAGGC	420
QY	421	ACCAGGGAAGTCTCCTGGGGCGATCTGTAAATAAACCTTTTCTTTTGTTTTAA	480
Db	421	ACCAGGGAAGTCTCCTGGGGCGATCTGTAAATAAACCTTTTCTTTTGTTTTAA	480
QY	481	AAAAAATAAAAGTCGACC	498
Db	481	AAAAAATAAAAGTCGACC	498
RESULT 2			
ID	AAA94623	standard; DNA; 1441 BP.	
XX	AAA94623;		
XX	AC		
DT	11-JAN-2001	(first entry)	
XX	Human CASB618	coding sequence.	
DE	Human CASB618	coding sequence.	
XX	Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian;		
KW	colon; autoimmune disease; HLA_A0201; ss.		
XX	Homo sapiens.		
OS	Homo sapiens.		
XX	Key	Location/Qualifiers	
FH	CDS	259..1221	
FT	FT	/*tag= a	
FT	FT	/product= "Human CASB618"	
PN	WO200053748-A2.		
XX	14-SEP-2000.		
PD	14-SEP-2000.		
XX	09-MAR-2000; 2000MO-EP002048.		
PF	09-MAR-2000; 2000MO-EP002048.		
XX	11-MAR-1999; 99GB-00005607.		
PR	01-SEP-1999; 99GB-00020590.		
XX	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.		
PA	Bruck CEM, Cassart J, Coche T, Vinals y De BassolsC;		
XX			

XX WPI; 2000-572268/53.
DR P-PSDB; AAB26325.
DR
XX
XX New human CASB618 polypeptide, useful as a vaccine for prophylactic and
PT therapeutic treatment of cancers, particularly ovarian or colon cancer,
PT autoimmune diseases and related conditions.
XX
XX
PS Claim 13; Page 61; 76pp; English.
XX
XX The present sequence is the coding sequence of human CASB618 protein. The
CC gene for human CASB618 is thought to be located on chromosome 15. The
CC protein encoded by the present sequence and epitopes of the CASB618
CC protein (see AAB26327 to AAB26399) are useful in diagnosing the occurrence
CC of tumour cells and in vaccines for prophylactic and therapeutic
CC treatment of cancers, particularly ovarian or colon cancer, autoimmune
CC diseases and related conditions
XX
XX Sequence 1441 BP; 289 A; 466 C; 392 G; 294 T; 0 U; 0 Other;
SQ

[illegible]

OS Homo sapiens.
XX WO200190304-A2.
XX 29-NOV-2001.
XX 18-MAY-2001; 2001WO-US016450.
XX 19-MAY-2000; 2000US-0205515P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Birse CE, Rosen CA;
XX WPI; 2002-122018/16.
XX P-PSDB; ABB90126.
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neutral, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders.
XX Claim 4; SEQ ID NO 1097; 2081pp + Sequence Listing; English.
XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABB9040-ABB9044) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1474 BP; 292 A; 480 C; 404 G; 296 T; 0 U; 2 Other;
Query Match 85.3%; Score 425; DB 6; Length 1474;
Best Local Similarity 100.0%; Pred. No. 1.1e-201;
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 62 GCGCCGCTTCTGGGTACACGCTGGCAACGCGCTCTGCTCTTCTCGAGAGGCGCG 121
DB 1024 GCGCCGCTTCTGGGTACACGCTGGCAACGCGCTCTGCTCTTCTCGAGAGGCGCG 1083
QY 122 TGGTGAATCTCCAGTATGTTGGCCACGCGCTCTTCTCGAGAGGCGCGCA 181
DB 1084 TGGTGAATCTCCAGTATGTTGGCCACGCGCTCTTCTCGAGAGGCGCGCA 1143
QY 182 AGGACTGCAGCCAGAGAGAGGCGGCTCACTTATCTCTGGCGACCACTGCACAGC 241
DB 1144 AGGACTGCAGCCAGAGAGAGGCGGCTCACTTATCTCTGGCGACCACTGCACAGC 1203
QY 242 AGGCGGCTCTCCAGACTTAATGTATACCACTAAGCTGTAGAGGAGGAGCCCAATCTGG 301
DB 1204 AGGCGGCTCTCCAGACTTAATGTATACCACTAAGCTGTAGAGGAGGAGCCCAATCTGG 1263
QY 302 ACTCCTTCCCGGCTTGGAGATCGCAGCGCGGAGAGAGTGGCCCGCAAGGCTGGGCCA 361
DB 1264 ACTCCTTCCCGGCTTGGAGATCGCAGCGCGGAGAGAGTGGCCCGCAAGGCTGGGCCA 1323
QY 362 GGAGAGCTCCAGAGAGGCGCACTGAGCGCTGTGGCGGAGGCGCTCGACATCCGAGGCA 421
DB 1324 GGAGAGCTCCAGAGAGGCGCACTGAGCGCTGTGGCGGAGGCGCTCGACATCCGAGGCA 1383
QY 422 CCAAGGAAAGTCTCTGGGGGATCTGTAAATAAACCTTTTCTTTGTTTAAAA 481

DB 1384 CCAGGAAAGTCTCTGGGGGATCTGTAAATAAACCTTTTCTTTGTTTAAAA 1443
QY 482 AAAA 486
DB 1444 AAAA 1448
RESULT 4
ADD19228
ID ADD19228 standard; cDNA; 1491 BP.
AC ADD19228;
XX 15-JAN-2004 (first entry)
DT 15-JAN-2004 (first entry)
XX Human cDNA from secreted protein gene 45.
DE human secreted protein; cytosolic; antibacterial; virucide;
XX neuroprotective; gynaecological; gastrointestinal-Gen; cardiac;
KW cardiovascular-Gen; nephrotropic; antiinflammatory; muscular-Gen;
KW respiratory-Gen; immunosuppressive; cerebroprotective; vasotropic;
KW neurotropic; antiallergic; cancer; bacterial infection; viral infection;
KW neural disorder; immune system disorder; blood disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW inflammatory disorder; proliferative disorder; human; ss; gene.
XX Homo sapiens.
XX WO2003052377-A2.
XX 26-JUN-2003.
XX 06-NOV-2002; 2002WO-US035606.
XX 07-NOV-2001; 2001US-0331046P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
XX WPI; 2003-533050/50.
XX P-PSDB; ADD19303.
XX New isolated nucleic acids encoding signal transduction pathway component
PT polypeptides, useful for diagnosing, treating, and/or preventing
PT disorders, such as cancer, infections, cardiovascular and inflammatory
PT diseases.
XX Claim 1; SEQ ID NO 55; 554pp; English.
XX The invention relates to an isolated nucleic acid molecule (cDNA)
CC encoding a human secreted protein, representing one of 85 novel genes.
CC Also included are recombinant vectors, host cells (expressing the
CC protein), the secreted proteins (including their fragments, epitopes and
CC homologues), an isolated antibody that binds specifically to the protein,
CC diagnosing a pathological condition or susceptibility to a pathological
CC condition (comprising determining the presence or absence of a mutation
CC in the nucleic acid and diagnosing a condition based on the presence or
CC absence of the mutation), diagnosing a pathological condition or
CC susceptibility to a pathological condition (comprising determining the
CC presence or amount of expression of the protein in a biological sample
CC and diagnosing a condition based on the presence or amount of expression
CC of the protein), preventing, treating or ameliorating a medical condition
CC by administering the nucleic acid or protein to a mammalian subject,
CC identifying a binding partner to the protein, the gene corresponding to
CC the cDNA sequence, and identifying an activity in a biological assay
CC (comprising expressing the nucleic acid in a cell, isolating the
CC supernatant, detecting an activity in a biological assay and identifying
CC the protein in the supernatant having the activity). The nucleic acids
CC and proteins display the following activities Cytostatic, antibacterial,
CC virucide, Neuroprotective, Gynaecological, Gastrointestinal-Gen,

CC Cardiant, Cardiovascular-Gen, Nephrotropic, Antiinflammatory, Muscular-
CC Gen, Respiratory-Gen, Immunosuppressive, Cerebroprotective, Vasotropic,
CC Nootropic, Antiallergic. The methods and compositions of the present
CC invention are useful for diagnosing, treating, preventing and/or
CC prognosticating disorders related to the novel polypeptides, such as
CC cancer, bacterial or viral infections, and neural, immune system, blood,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal, inflammatory or proliferative disorders (many examples of these
CC diseases and disorders are given in the specification). The present
CC sequence encodes a novel secreted protein of the invention.
XX
SQ Sequence 1491 BP; 293 A; 485 C; 410 G; 302 T; 0 U; 1 Other;

Query Match 85.3%; Score 425; DB 9; Length 1491;
Best Local Similarity 100.0%; Pred. No. 1.1e-201;
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 GCGCCGCTTCTGGGTACGCTGGCAACCGCGTCTGTGCTCTTCTCGAGGGCGCG 121
DB 1041 GCGCCGCTTCTGGGTACGCTGGCAACCGCGTCTGTGCTCTTCTCGAGGGCGCG 1100
QY 122 TGGTGAAGTCTCCAGTATGTTGGCCGCGCTCTTGGCACCCCTTCTGACCAAGCGCCA 181
DB 1101 TGGTGAAGTCTCCAGTATGTTGGCCGCGCTCTTGGCACCCCTTCTGACCAAGCGCCA 1160
QY 182 AGGACTGACGACGAGAGAGAGGGGCTCACTCTTATCTCTGGCGGACCCACTGACCAAGC 241
DB 1161 AGGACTGACGACGAGAGAGAGGGGCTCACTCTTATCTCTGGCGGACCCACTGACCAAGC 1220
QY 242 AGGCGGCTCTCCAGACTTAATATGATACCACTTAACCTGTGAGGGGGAGCCCAATCTGG 301
DB 1221 AGGCGGCTCTCCAGACTTAATATGATACCACTTAACCTGTGAGGGGGAGCCCAATCTGG 1280
QY 302 ACTCCTTCCCGCGCTTGGGACATCGCAGCCCGGGAAGCAAGTCCCGCCAGGCTGGGCCA 361
DB 1281 ACTCCTTCCCGCGCTTGGGACATCGCAGCCCGGGAAGCAAGTCCCGCCAGGCTGGGCCA 1340
QY 362 GGAAGAGCTCCAGAGAGGAGCACTGAGCGCTGCTGGCGCGAGGCGCTGGACATCCGACAGCA 421
DB 1341 GGAAGAGCTCCAGAGAGGAGCACTGAGCGCTGCTGGCGCGAGGCGCTGGACATCCGACAGCA 1400
QY 422 CCAGGGAAGTCTCTCTGGGCGGATCTGTAATAATAAACCCTTTTCTTTTAAAA 481
DB 1401 CCAGGGAAGTCTCTCTGGGCGGATCTGTAATAATAAACCCTTTTCTTTTAAAA 1460
QY 482 AAAAA 486
DB 1461 AAAAA 1465

RESULT 5
AAA96505
ID AAA96505 standard; cDNA; 1421 BP.
XX
AC AAA96505;
XX
DT 08-FEB-2001 (first entry)
XX
DE cDNA encoding a human transmembrane protein.
XX
KW Human; transmembrane protein; cell proliferation disorder; myeloma;
KW reproductive disorder; smooth muscle disorder; neurological disorder;
KW arteriosclerosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
KW allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;
KW Alzheimer's disease; Tourette's disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 267..1229
FT /*tag= a
XX
PN WO200056891-A2.

XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US007817.
XX
PR 22-MAR-1999; 99US-0125537P.
PR 16-JUN-1999; 99US-0139565P.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Yue H, Lal P, Tang YT, Hillman JL, Reddy R, Bandman O;
PI Baughn MR, Lu DAM, Azimzai Y, Yang J;
XX
DR WPI; 2000-579485/54.
DR P-PSDB; AAB18992.
XX

PT New human transmembrane proteins are used to treat a disease or condition
PT associated with decreased expression of functional HTMP e.g. Tourette's
PT disorder, angina and leukemia.
PS
PS Claim 4; Page 129; 130pp; English.

XX
CC The present sequence encodes a human transmembrane proteins (HTMP).
CC Agonists and antagonists of the protein are used to treat a disease or
CC condition associated with overexpression of the protein. Diseases and
CC conditions which can be treated include cell proliferative,
CC immunological, reproductive, smooth muscle and neurological disorders
CC e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency
CC syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,
CC stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The
CC polynucleotides may be used to detect and quantify gene expression in
CC biopsied tissues where protein expression may be correlated with disease
CC e.g. to determine absence, presence or excess expression of HTMP or to
CC monitor regulation of HTMP expression during therapeutic intervention
XX
SQ Sequence 1421 BP; 264 A; 468 C; 395 G; 294 T; 0 U; 0 Other;

Query Match 83.5%; Score 416; DB 3; Length 1421;
Best Local Similarity 100.0%; Pred. No. 3.4e-197;
Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 GCGCCGCTTCTGGGTACGCTGGCAACCGCGTCTGTGCTCTTCTCGAGGGCGCG 121
DB 1006 GCGCCGCTTCTGGGTACGCTGGCAACCGCGTCTGTGCTCTTCTCGAGGGCGCG 1065
QY 122 TGGTGAAGTCTCCAGTATGTTGGCCGCGCTCTTGGCACCCCTTCTGACCAAGCGCCA 181
DB 1066 TGGTGAAGTCTCCAGTATGTTGGCCGCGCTCTTGGCACCCCTTCTGACCAAGCGCCA 1125
QY 182 AGGACTGACGACGAGAGAGAGGGGCTCACTCTTATCTCTGGCGGACCCACTGACAGCA 241
DB 1126 AGGACTGACGACGAGAGAGAGGGGCTCACTCTTATCTCTGGCGGACCCACTGACAGCA 1185
QY 242 AGGCGGCTCTCCAGACTTAATATGATACCACTTAACCTGTGAGGGGGAGCCCAATCTGG 301
DB 1186 AGGCGGCTCTCCAGACTTAATATGATACCACTTAACCTGTGAGGGGGAGCCCAATCTGG 1245
QY 302 ACTCCTTCCCGCGCTTGGGACATCGCAGGCGGGAAGCAAGTCCCGCCAGGCTGGGCCA 361
DB 1246 ACTCCTTCCCGCGCTTGGGACATCGCAGGCGGGAAGCAAGTCCCGCCAGGCTGGGCCA 1305
QY 362 GGAAGAGCTCCAGAGAGGAGCACTGAGCGCTGCTGGCGGAGGCTCGGACATCCGACAGCA 421
DB 1306 GGAAGAGCTCCAGAGAGGAGCACTGAGCGCTGCTGGCGGAGGCTCGGACATCCGACAGCA 1365
QY 422 CCAGGGAAGTCTCTCTGGGCGGATCTGTAATAATAAACCCTTTTCTTTTGT 477
DB 1366 CCAGGGAAGTCTCTCTGGGCGGATCTGTAATAATAAACCCTTTTCTTTTGT 1421

RESULT 6
ADA53627/c
ID ADA53627 standard; cDNA; 2684 BP.

```
XX AC ADA53627;
XX DT 20-NOV-2003 (first entry)
XX DE Human coding sequence, SEQ ID 1195.
XX KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
XX KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
XX KW inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX OS Homo sapiens.
XX PN EPI293569-A2.
XX PD 19-MAR-2003.
XX PF 21-MAR-2002; 2002EP-00006586.
XX PR 14-SEP-2001; 2001JP-00328381.
XX PR 24-JAN-2002; 2002US-0350435P.
XX PA (HELI-) HELIX RES INST.
XX PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
XX PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
XX PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
XX DR WPI; 2003-395539/38.
XX DR P-PSDB; ADA55266.
XX PT New polynucleotides encoding full-length polypeptides, e.g. secretory
XX PT and/or membrane proteins, useful for developing medicines for diseases in
XX PT which the gene is involved, or as target molecules for gene therapy.
XX PS Claim 1; SEQ ID NO 1195; 205pp; English.
XX CC The present invention relates to novel human secretory or membrane
XX CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
XX CC ADA54071). The coding sequences are useful in the gene therapy of
XX CC diseases caused by abnormalities of the proteins, e.g. cancer,
XX CC inflammatory diseases, osteoporosis or neurological disease.
XX SQ Sequence 2684 BP; 599 A; 785 C; 684 G; 616 T; 0 U; 0 Other;

Query Match      78.7%; Score 392; DB 7; Length 2684;
Best Local Similarity 100.0%; Pred. No. 3.3e-185;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GGGCTCTGTGCTCTTCTCTGGAGGGGGCGGTGTGAGTCTCCAGTATGTTGGCCACG 150
DB 2343 GGGCTCTGTGCTCTTCTCTGGAGGGGGCGGTGTGAGTCTCCAGTATGTTGGCCACG 2284
QY 151 GCTCTTCGACCCCTTCTGGACCAAGCGCCAGAGTGCAGCCAGAGAGAGGGGCTCA 210
DB 2283 GCTCTTCGACCCCTTCTGGACCAAGCGCCAGAGTGCAGCCAGAGAGAGGGGCTCA 2224
QY 211 CCTCTTATCCTCGGCGACCCACTGCACAAGAGGCGCTCTCCAGACTTAATATATC 270
DB 2223 CCTCTTATCCTCGGCGACCCACTGCACAAGAGGCGCTCTCCAGACTTAATATATC 2164
QY 271 ACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCCTTGGACATCGAGG 330
DB 2163 ACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCCTTGGACATCGAGG 2104
QY 331 CCGGGAAGCAGTGGCCCGCAGGCTGGCCAGAGAGCTCCAGAAAGGCACTGAGCGCT 390
DB 2103 CCGGGAAGCAGTGGCCCGCAGGCTGGCCAGAGAGCTCCAGAAAGGCACTGAGCGCT 2044
QY 391 GCTGGCGGAGGCTCGGACATCCGAGGACCCAGGAAAGTCTCTGGGCGATCTGTGA 450
DB 2043 GCTGGCGGAGGCTCGGACATCCGAGGACCCAGGAAAGTCTCTCTGGGCGATCTGTGA 1984
```

```
QY 451 AATAACCTTTTCTTCTTTGTTTAAATA 482
DB 1983 AATAACCTTTTCTTCTTTGTTTAAATA 1952

RESULT 7
ABZ68115
ID ABZ68115 standard; DNA; 5033 BP.
XX ABZ68115;
AC ABZ68115;
XX
DT 26-MAR-2003 (first entry)
XX
DE Human secreted protein encoding genomic DNA SEQ ID NO 1638.
XX
KW Human; secreted protein; nootropic; neuroprotective; cytostatic;
KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW vulnery; antibacterial; antiparkinsonian; antislaking; antihaemic;
KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
KW antiinflammatory; antiallergic; antidiabetic; antilucer; anticonvulsant;
KW antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine;
KW cardiovascular disorder; neurological disease; nephrotropic;
KW gene therapy; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277186-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-US009188.
XX
PR 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-040583/03.
XX
XX
XX PT New human secreted proteins encoded by genes contained in cDNA clones
XX PT (e.g. HGCAC19), useful for preventing, treating or diagnosing e.g. AIDS,
XX PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
XX PT West Nile fever.
XX PS Disclosure; Page 2263-2264; 2423pp; English.
XX
CC The invention relates to novel human genes (ABZ68891-ABZ68209) and the
CC encoded secreted proteins (ABP9470-ABP9872) useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections
XX
SQ Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;

Query Match      78.7%; Score 392; DB 7; Length 5033;
Best Local Similarity 100.0%; Pred. No. 3.4e-185;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GGGCTCTGTGCTCTTCTCTGGAGGGGGCGGTGTGAGTCTCCAGTATGTTGGCCACG 150
```

```
Db 3380 GGGCTCTGTGCTCTTCTCGAGGGGCGGTGTGAGTCTCCAGTATGTTCCGCCAGC 3439
QY 151 GCTCTTGACACCTTCTTGACCAAGGCCCAAGACTGCAGCCAGAGAGAGGGGCTCA 210
Db 3440 GCTCTTGACACCTTCTTGACCAAGGCCCAAGACTGCAGCCAGAGAGAGGGGCTCA 3499
QY 211 CCTTTATCTCTGGCGACCCACTGCACAGAGGCGCTCTCCAGACTTAAATGTATC 270
Db 3500 CCTTTATCTCTGGCGACCCACTGCACAGAGGCGCTCTCCAGACTTAAATGTATC 3559
QY 271 ACCACTAACCCTGTAGGGGGACCCAACTGTGACTCTCCCGCTTGGACATCGCAGG 330
Db 3560 ACCACTAACCCTGTAGGGGGACCCAACTGTGACTCTCCCGCTTGGACATCGCAGG 3619
QY 331 CCGGAAGCAGTGGCCCGCAGGCGCTGGCCAGAGAGCTCCAGGAAGGCACTGAGCGCT 390
Db 3620 CCGGAAGCAGTGGCCCGCAGGCGCTGGCCAGAGAGCTCCAGGAAGGCACTGAGCGCT 3679
QY 391 GCTGGCGGAGGCGCTCGGACATCCGAGGACCCAGGAAAGTCTCTGGGCGATCTGTA 450
Db 3680 GCTGGCGGAGGCGCTCGGACATCCGAGGACCCAGGAAAGTCTCTGGGCGATCTGTA 3739
QY 451 AATAAACCTTTTCTTTTGTGTTTAAAAA 482
Db 3740 AATAAACCTTTTCTTTTGTGTTTAAAAA 3771

RESULT 8
ABZ74587
ID ABZ74587 standard; DNA; 5033 BP.
XX
AC ABZ74587;
XX
DT 12-MAY-2003 (first entry)
XX
DE Secreted protein gene 360 genomic fragment HUFCL31, SEQ ID NO:1734.
XX
KW Human; secreted protein; cancer; tumour; hyperproliferative disorder;
KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;
KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
KW drug screening; chromosome identification; chromosome mapping;
KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
KW antianaemic; vulnery; chromosome 9p21; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277013-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-US009370.
XX
PR 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-040578/03.
XX
PT New human secreted proteins and nucleic acids, useful for detecting or
PT treating cancer or other hyperproliferative disorders, autoimmune
PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
XX
PS Disclosure; Page 2315-2316; 2474pp; English.
XX
CC ABZ73281-ABZ73697 represent CDNAs corresponding to 391 human secreted
CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
CC invention also encompasses antibodies specific for the secreted proteins,
```

```
CC the use of the secreted proteins in drug screening and recombinant
CC vectors and host cells comprising a nucleic acid of the invention. The
CC secreted proteins are thought to be involved in biological activities
CC associated with cellular signalling, cellular differentiation, cell
CC migration, prohormone activation and neurotransmitter activity. The
CC secreted proteins, nucleic acids encoding them, antibodies or antibody
CC fragments specific for the secreted proteins, and modulators of protein
CC activity are useful for diagnosing or treating cancers or other
CC hyperproliferative disorders. Additionally, the secreted proteins and
CC their nucleic acids may also be used in the treatment of autoimmune
CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
CC wound healing. Nucleic acids of the invention may be used for chromosome
CC identification, chromosome mapping, in gene therapy, for identifying
CC individuals from minute biological samples, as hybridisation probes, and
CC as molecular weight markers. The present sequence represents a human
CC secreted protein genomic fragment referred to in the disclosure of the
CC invention
XX
SQ Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;
Query Match 78.7%; Score 392; DB 7; Length 5033;
Best Local Similarity 100.0%; Pred. No. 3.4e-185;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 91 GGGCTCTGTGCTCTTCTCTCGAGGGGCGGTGTGAGTCTCCAGTATGTTCCGCCAGC 150
Db 3380 GGGCTCTGTGCTCTTCTCTCGAGGGGCGGTGTGAGTCTCCAGTATGTTCCGCCAGC 3439
QY 151 GCTCTTGACACCTTCTTGACCAAGGCCCAAGACTGCAGCCAGAGAGAGGGGCTCA 210
Db 3440 GCTCTTGACACCTTCTTGACCAAGGCCCAAGACTGCAGCCAGAGAGAGGGGCTCA 3499
QY 211 CCTTTATCTCTGGCGACCCACTGCACAGAGGCGCTCTCCAGACTTAAATGTATC 270
Db 3500 CCTTTATCTCTGGCGACCCACTGCACAGAGGCGCTCTCCAGACTTAAATGTATC 3559
QY 271 ACCACTAACCCTGTAGGGGGACCCAACTGTGACTCTCCCGCTTGGACATCGCAGG 330
Db 3560 ACCACTAACCCTGTAGGGGGACCCAACTGTGACTCTCCCGCTTGGACATCGCAGG 3619
QY 331 CCGGAAGCAGTGGCCCGCAGGCGCTGGCCAGAGAGCTCCAGGAAGGCACTGAGCGCT 390
Db 3620 CCGGAAGCAGTGGCCCGCAGGCGCTGGCCAGAGAGCTCCAGGAAGGCACTGAGCGCT 3679
QY 391 GCTGGCGGAGGCGCTCGGACATCCGAGGACCCAGGAAAGTCTCTGGGCGATCTGTA 450
Db 3680 GCTGGCGGAGGCGCTCGGACATCCGAGGACCCAGGAAAGTCTCTGGGCGATCTGTA 3739
QY 451 AATAAACCTTTTCTTTTGTGTTTAAAAA 482
Db 3740 AATAAACCTTTTCTTTTGTGTTTAAAAA 3771

RESULT 9
ADC21005
ID ADC21005 standard; DNA; 5033 BP.
XX
AC ADC21005;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human secreted protein-related DNA sequence #423.
XX
KW gene therapy; human; secreted protein; haemopoietic disorder;
KW haematological disorder; anaemia; haemophilia; inflammatory disorder;
KW inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
KW leukaemia; wound healing; epithelial cell proliferation disorder;
KW immune disorder; autoimmune disorder; asthmatic disorder;
KW cardiovascular disorder; atherosclerosis; myocarditis;
KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;
KW gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.
```

OS Homo sapiens.
XX
PN WO200292787-A2.
XX
PD 21-NOV-2002.
XX
PF 26-MAR-2002; 2002WO-US009257.
XX
PR 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-129287/12.
XX
PT New human secreted proteins and nucleic acid molecules, useful for
PT preparing a diagnostic or pharmaceutical composition for diagnosing,
PT preventing or treating hematopoietic or hematologic disorders, e.g.
PT anemia or hemophilia.
XX
PS Disclosure; SEQ ID NO 959; 1512pp; English.
XX
CC The invention comprises the amino acid and coding sequences of human
CC secreted proteins. The DNA and protein sequences of the invention are
CC useful for detecting, preventing, diagnosing, prognosticating, treating
CC or ameliorating: hematopoietic or hematological disorders (e.g. anaemia
CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease
CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
CC wound healing and disorders of epithelial cell proliferation; immune
CC disorders (e.g. autoimmune disorders and asthmatic disorders);
CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);
CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
CC and gastrointestinal disorders (e.g. duodenal ulcers and
CC gastroenteritis). The present DNA sequence was used in the
XX
SQ Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;

Query Match 78.7%; Score 392; DB 9; Length 5033;
Best Local Similarity 100.0%; Pred. No. 3.4e-185;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 91 GGGGTCCTGTGCTCTTCTCTGAGAGGGGCGGTGTGAGTCTCCAGTATGTTGCGCCGAC 150
Db 3380 GGGGTCCTGTGCTCTTCTCTGAGAGGGGCGGTGTGAGTCTCCAGTATGTTGCGCCGAC 3439
QY 151 GCTCTTGCACCCCTTCTTGACCAAGCGCCAGGAGTGCAGCCAGAGAGAGGGGCTCA 210
Db 3440 GCTCTTGCACCCCTTCTTGACCAAGCGCCAGGAGTGCAGCCAGAGAGAGGGGCTCA 3499
QY 211 CCTCTTATCCTGGGCGACCCACTGCACAGAGGCGCTCTCCAGACTTAAATGTATC 270
Db 3500 CCTCTTATCCTGGGCGACCCACTGCACAGAGGCGCTCTCCAGACTTAAATGTATC 3559
QY 271 ACCACTAACCTGTGAGGGGGAGCCCAATCTGACTCTTCCCCCGCTTGGGACATCGCAGG 330
Db 3560 ACCACTAACCTGTGAGGGGGAGCCCAATCTGACTCTTCCCCCGCTTGGGACATCGCAGG 3619
QY 331 CCGGGAAGCAGTGGCCCGCAGGCGCTGGGCGCAGAGAGCTCCAGAAAGGCACTGAGCGCT 390
Db 3620 CCGGGAAGCAGTGGCCCGCAGGCGCTGGGCGCAGAGAGCTCCAGAAAGGCACTGAGCGCT 3679
QY 391 GCTGGCGCGAGGCTCGGACATCCGACGACCAAGGAAAGTCTCTGGGCGGATCTGTA 450
Db 3680 GCTGGCGCGAGGCTCGGACATCCGACGACCAAGGAAAGTCTCTGGGCGGATCTGTA 3739
QY 451 AATAAACCTTTTCTTTCTTTTAAAAA 482
Db 3740 AATAAACCTTTTCTTTCTTTTAAAAA 3771

RESULT 10
AAC95564
ID AAC95564 standard; cDNA; 1460 BP.
XX
AC AAC95564;
XX
DT 21-FEB-2001 (first entry)
XX
DE Human secreted protein gene 44 SEQ ID NO:54.
XX
KW Human; secreted protein; cyostatic; immunosuppressive; nootropic;
KW neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antitumor; anticonvulsant; antibacterial;
KW antifungal; antiparasitic; cardiant; cancer; immune disease; allergy;
KW cardiovascular disorder; wound healing; infection; neurological disease;
KW ss.
XX
OS Homo sapiens.
XX
PN WO200061596-A1.
XX
PD 19-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US008983.
XX
PR 09-APR-1999; 99US-0128703P.
PR 14-JAN-2000; 2000US-0176068P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-611865/58.
DR P-PSDB; AAB52055.
XX
PT Fifty nucleic acid molecules encoding human secreted proteins, useful in
PT the prevention, treatment and diagnosis of cancer, immune disorders,
PT cardiovascular disorders and neurological diseases.
XX
PS Claim 1; Page 443; 505pp; English.

QY 91 GGGGTCCTGTGCTCTTCTCTGAGAGGGGCGGTGTGAGTCTCCAGTATGTTGCGCCGAC 150
Db 3380 GGGGTCCTGTGCTCTTCTCTGAGAGGGGCGGTGTGAGTCTCCAGTATGTTGCGCCGAC 3439
QY 151 GCTCTTGCACCCCTTCTTGACCAAGCGCCAGGAGTGCAGCCAGAGAGAGGGGCTCA 210
Db 3440 GCTCTTGCACCCCTTCTTGACCAAGCGCCAGGAGTGCAGCCAGAGAGAGGGGCTCA 3499
QY 211 CCTCTTATCCTGGGCGACCCACTGCACAGAGGCGCTCTCCAGACTTAAATGTATC 270
Db 3500 CCTCTTATCCTGGGCGACCCACTGCACAGAGGCGCTCTCCAGACTTAAATGTATC 3559
QY 271 ACCACTAACCTGTGAGGGGGAGCCCAATCTGACTCTTCCCCCGCTTGGGACATCGCAGG 330
Db 3560 ACCACTAACCTGTGAGGGGGAGCCCAATCTGACTCTTCCCCCGCTTGGGACATCGCAGG 3619
QY 331 CCGGGAAGCAGTGGCCCGCAGGCGCTGGGCGCAGAGAGCTCCAGAAAGGCACTGAGCGCT 390
Db 3620 CCGGGAAGCAGTGGCCCGCAGGCGCTGGGCGCAGAGAGCTCCAGAAAGGCACTGAGCGCT 3679
QY 391 GCTGGCGCGAGGCTCGGACATCCGACGACCAAGGAAAGTCTCTGGGCGGATCTGTA 450
Db 3680 GCTGGCGCGAGGCTCGGACATCCGACGACCAAGGAAAGTCTCTGGGCGGATCTGTA 3739
QY 451 AATAAACCTTTTCTTTCTTTTAAAAA 482
Db 3740 AATAAACCTTTTCTTTCTTTTAAAAA 3771

QY 259 TTTAAATGTATCACCCTAACCTGTGAGGGGACCCCAATCTGACTCCTTCCCGCCTTG 318
 DB 1217 TTTAAATGTATCACCCTAACCTGTGAGGGGACCCCAATCTGACTCCTTCCCGCCTTG 1276
 QY 319 GGACATCGCAGGCGCGGGAAGCAGTGTCCCGCCAGGCTGGGCCAGAGAGCTCCAGGAAG 378
 DB 1277 GGACATCGCAGGCGCGGGAAGCAGTGTCCCGCCAGGCTGGGCCAGAGAGCTCCAGGAAG 1336
 QY 379 GCACCTGAGCGCTGTGCGCGGAGGCTCGACATCCGACAGGACCAAGGAAGTCTCTG 438
 DB 1337 GCACCTGAGCGCTGTGCGCGGAGGCTCGACATCCGACAGGACCAAGGAAGTCTCTG 1396
 QY 439 GGGCGATCTGTAATAAACCCTTTTCTTTTCTTTTAAAAAAA 486
 DB 1397 GGGCGATCTGTAATAAACCCTTTTCTTTTCTTTTAAAAAAA 1444

RESULT 11

ABZ67235

ID ABZ67235 standard; cDNA; 1460 BP.

AC ABZ67235;

DT 26-MAR-2003 (first entry)

DE Human secreted protein encoding cDNA SEQ ID NO 355.

XX Human; secreted protein; nootropic; neuroprotective; cytosstatic;
 KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
 KW vulnerary; antibacterial; antiparkinsonian; antischling; antianaemic;
 KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
 KW antiinflammatory; antiallergic; antidiabetic; antidiacer; anticonvulsant;
 KW antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine;
 KW cardiovascular disorder; neurological disease; nephrotropic;
 KW gene therapy; gene; chromosome 9p21; ds.

OS Homo sapiens.

PN WO200277186-A2.

PD 03-OCT-2002.

PF 26-MAR-2002; 2002WO-US009188.

PR 27-MAR-2001; 2001US-0278650P.

PR 12-SEP-2001; 2001US-00950082.

PR 12-SEP-2001; 2001US-00950083.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR WPI: 2003-040583/03.

DR P-PSDB; ABP99814.

PT New human secreted proteins encoded by genes contained in cDNA clones
 PT (e.g. HGCAC19), useful for preventing, treating or diagnosing e.g. AIDS,
 PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
 PT West Nile fever.

PS Claim 7; Page 1353; 2423pp; English.

CC The invention relates to novel human genes (ABZ66891-ABZ68209) and the
 CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
 CC treating or ameliorating medical conditions e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid

CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
 CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections

XX Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;

Query Match 45.8%; Score 228; DB 7; Length 1460;

Best Local Similarity 100.0%; Pred. No. 2.2e-103; Indels 0; Gaps 0;

Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 TTTAAATGTATCACCCTAACCTGTGAGGGGACCCCAATCTGACTCCTTCCCGCCTTG 318
 DB 1217 TTTAAATGTATCACCCTAACCTGTGAGGGGACCCCAATCTGACTCCTTCCCGCCTTG 1276
 QY 319 GGACATCGCAGGCGCGGGAAGCAGTGTCCCGCCAGGCTGGGCCAGAGAGCTCCAGGAAG 378
 DB 1277 GGACATCGCAGGCGCGGGAAGCAGTGTCCCGCCAGGCTGGGCCAGAGAGCTCCAGGAAG 1336
 QY 379 GCACCTGAGCGCTGTGCGCGGAGGCTCGACATCCGACAGGACCAAGGAAGTCTCTG 438
 DB 1337 GCACCTGAGCGCTGTGCGCGGAGGCTCGACATCCGACAGGACCAAGGAAGTCTCTG 1396
 QY 439 GGGCGATCTGTAATAAACCCTTTTCTTTTCTTTTAAAAAAA 486
 DB 1397 GGGCGATCTGTAATAAACCCTTTTCTTTTCTTTTAAAAAAA 1444

RESULT 12

ABZ73640

ID ABZ73640 standard; cDNA; 1460 BP.

AC ABZ73640;

DT 12-MAY-2003 (first entry)

DE Secreted protein-encoding gene 360 cDNA clone HUFCL31, SEQ ID NO:370.

XX Human; secreted protein; cancer; tumour; hyperproliferative disorder;
 KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW drug screening; chromosome identification; chromosome mapping;
 KW cytosstatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
 KW antianaemic; vulnerary; chromosome 9p21; gene; ss.

OS Homo sapiens.

PN WO200277013-A2.

PD 03-OCT-2002.

PF 26-MAR-2002; 2002WO-US009370.

PR 27-MAR-2001; 2001US-0278650P.

PR 12-SEP-2001; 2001US-00950082.

PR 12-SEP-2001; 2001US-00950083.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR WPI: 2003-040578/03.

DR P-PSDB; ABR01306.

PT New human secreted proteins and nucleic acids, useful for detecting or
 PT treating cancer or other hyperproliferative disorders, autoimmune
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
 XX Claim 21; Page 1345; 2474pp; English.

CC ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
 CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
 CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The

CC invention also encompasses antibodies specific for the secreted proteins,
CC the use of the secreted proteins in drug screening and recombinant
CC vectors and host cells comprising a nucleic acid of the invention. The
CC secreted proteins are thought to be involved in biological activities
CC associated with cellular signalling, cellular differentiation, cell
CC migration, prohormone activation and neurotransmitter activity. The
CC secreted proteins, nucleic acids encoding them, antibodies or antibody
CC fragments specific for the secreted proteins, and modulators of protein
CC activity are useful for diagnosing or treating cancers or other
CC hyperproliferative disorders. Additionally, the secreted proteins and
CC their nucleic acids may also be used in the treatment of autoimmune
CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
CC wound healing. Nucleic acids of the invention may be used for chromosome
CC identification, chromosome mapping, in gene therapy, for identifying
CC individuals from minute biological samples, as hybridisation probes, and
CC as molecular weight markers. The present sequence represents a human
CC secreted protein-encoding cDNA clone of the invention

XX SQ Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;

Query Match 45.8%; Score 228; DB 7; Length 1460;
Best Local Similarity 100.0%; Pred. No. 2.2e-103;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 TTAATAATGTATACCACTAAGCTGTAGGGGAGACCCCAATCTGACTCTCCCGCCTTG 318
DB 1217 TTAATAATGTATACCACTAAGCTGTAGGGGAGACCCCAATCTGACTCTCCCGCCTTG 1276

QY 319 GGACATCGCAGGCGGGAGAGAGTGCCTCCAGGCTCCAGAGAGTCCAGAGAGG 378
DB 1277 GGACATCGCAGGCGGGAGAGAGTGCCTCCAGGCTCCAGAGAGTCCAGAGAGG 1336

QY 379 GCACGTAGCGCTGCTGGCGGAGGCTCGGACATCCGACAGCACCAGGAGAAAGTCTCTG 438
DB 1337 GCACGTAGCGCTGCTGGCGGAGGCTCGGACATCCGACAGCACCAGGAGAAAGTCTCTG 1396

QY 439 GGGCGATCTGTAAATAAACCTTTTCTTTGTTTAAAAA 486
DB 1397 GGGCGATCTGTAAATAAACCTTTTCTTTGTTTAAAAA 1444

RESULT 13
ADC20289
ID ADC20289 standard; DNA; 1460 BP.

XX ADC20289;

AC 18-DEC-2003 (first entry)

DT Human secreted protein coding sequence #228.

DE Human secreted protein coding sequence #228.

XX gene therapy; human; secreted protein; haemopoietic disorder;
KW haematological disorder; anaemia; haemophilia; inflammatory disorder;
KW inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
KW leukaemia; wound healing; epithelial cell proliferation disorder;
KW immune disorder; autoimmune disorder; asthmatic disorder;
KW cardiovascular disorder; atherosclerosis; myocarditis;
KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;
KW gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.

OS Homo sapiens.

XX WO200292787-A2.

PN 21-NOV-2002.

PD 26-MAR-2002; 2002WO-US009257.

PF 27-MAR-2001; 2001US-0278650P.

PR 12-SEP-2001; 2001US-00950082.

PR 12-SEP-2001; 2001US-00950083.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM;

XX DR WPI; 2003-129287/12.

XX PT New human secreted proteins and nucleic acid molecules, useful for
PT preparing a diagnostic or pharmaceutical composition for diagnosing,
PT preventing or treating hematopoietic or hematologic disorders, e.g.
PT anemia or hemophilia.

XX PS Claim 1; SEQ ID NO 238; 1512bp; English.

XX The invention comprises the amino acid and coding sequences of human
CC secreted proteins. The DNA and protein sequences of the invention are
CC useful for detecting, preventing, diagnosing, prognosticating, treating
CC or ameliorating: haematopoietic or haematological disorders (e.g. anaemia
CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease
CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
CC wound healing and disorders of epithelial cell proliferation; immune
CC disorders (e.g. autoimmune disorders and asthmatic disorders);
CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);
CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
CC and gastrointestinal disorders (e.g. duodenal ulcers and
CC gastroenteritis). The present DNA sequence encodes a human secreted
CC protein of the invention.

XX SQ Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;

Query Match 45.8%; Score 228; DB 9; Length 1460;
Best Local Similarity 100.0%; Pred. No. 2.2e-103;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 TTAATAATGTATACCACTAAGCTGTAGGGGAGACCCCAATCTGACTCTCCCGCCTTG 318
DB 1217 TTAATAATGTATACCACTAAGCTGTAGGGGAGACCCCAATCTGACTCTCCCGCCTTG 1276

QY 319 GGACATCGCAGGCGGGAGAGAGTGCCTCCAGGCTCCAGAGAGTCCAGAGAGG 378
DB 1277 GGACATCGCAGGCGGGAGAGAGTGCCTCCAGGCTCCAGAGAGTCCAGAGAGG 1336

QY 379 GCACGTAGCGCTGCTGGCGGAGGCTCGGACATCCGACAGCACCAGGAGAAAGTCTCTG 438
DB 1337 GCACGTAGCGCTGCTGGCGGAGGCTCGGACATCCGACAGCACCAGGAGAAAGTCTCTG 1396

QY 439 GGGCGATCTGTAAATAAACCTTTTCTTTGTTTAAAAA 486
DB 1397 GGGCGATCTGTAAATAAACCTTTTCTTTGTTTAAAAA 1444

RESULT 14
AAK81950
ID AAK81950 standard; DNA; 525 BP.

XX AAK81950;

AC 07-NOV-2001 (first entry)

DT Human immune/haematopoietic antigen genomic sequence SEQ ID NO:36762.

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:36762.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.

OS Homo sapiens.

XX WO200157182-A2.

PN 09-AUG-2001.

PD 17-JAN-2001; 2001WO-US001354.

PF 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

XX

PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.

PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI, 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and metastasis.

Disclosure; SEQ ID NO 36762; 3071bp + Sequence Listing; English.

CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 525 BP; 174 A; 76 C; 146 G; 129 T; 0 U; 0 Other;

Query Match 5.2%; Score 26; DB 4; Length 525;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 464 TTCTTTTGTGTTTAAAAAAATAA 469
|||||
DQ 500 TTCTTTTGTGTTTAAAAAAATAA 525

RESULT 15
AAK81951
ID AAK81951 standard; DNA; 525 BP.
XX
AC AAK81951;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:36763.
XX
KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KM cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0188874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.

Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 464 TTCTTTGTTTTTTAAAAAAATAA 489
Db 500 TTCTTTGTTTTTTAAAAAAATAA 525

Search completed: February 23, 2004, 23:49:22
Job time : 204.357 secs

PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM,
XX
DR WPI; 2001-483426/52.
XX
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX
PS Disclosure; SEQ ID NO 36763; 3071pp + Sequence Listing; English.
XX
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 525 BP; 174 A; 76 C; 146 G; 129 T; 0 U; 0 Other;

Query Match 5.2%; Score 26; DB 4; Length 525;

PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 1143
SOFTWARE: PERL Program
SEQ ID NO 937
LENGTH: 1492
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No. 6673549 3257507CB1
US-09-976-594-937

Query Match 4.2%; Score 21; DB 4; Length 1492;
Best Local Similarity 100.0%; Pred. No. 0.83;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 TTTGTTTAAAAAAAT 487
DB 1394 TTTGTTTAAAAAAAT 1414

RESULT 3

US-07-705-490-1
Sequence 1, Application US/07705490
Patent No. 6107025
GENERAL INFORMATION:
APPLICANT: Caskey, C. T.
APPLICANT: Nelson, David L.
APPLICANT: Pieretti, Maura
APPLICANT: Warren, Stephen T.
APPLICANT: Oostra, Ben A.
TITLE OF INVENTION: Diagnosis of the Fragile X Syndrome
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thomas D. Paul
STREET: 1301 McKinney, Suite 5100
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77010-3095
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/705,490
FILING DATE: 19910708
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Paul, Thomas D.
REGISTRATION NUMBER: 32,714
REFERENCE/DOCKET NUMBER: D-5350
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713/651-5325
TELEFAX: 713/651-5246
TELEX: 762829
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3765 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-07-705-490-1

Query Match 4.0%; Score 20; DB 3; Length 3765;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 458 CTTTCTTTCTTTGTTT 477

DB 3159 CTTTCTTTCTTTGTTT 3178

RESULT 4

US-07-751-891B-1
Sequence 1, Application US/07751891B
Patent No. 6180337
GENERAL INFORMATION:
APPLICANT: Caskey, C. T.
APPLICANT: Nelson, David L.
APPLICANT: Pieretti, Maura
APPLICANT: Warren, Stephen T.
APPLICANT: Oostra, Ben A.
APPLICANT: Fu, Ying-hui
TITLE OF INVENTION: Diagnosis of the Fragile X Syndrome
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thomas D. Paul
STREET: 1301 McKinney, Suite 5100
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77010-3095
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/751,891B
FILING DATE: 29-Aug-1991
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Paul, Thomas D.
REGISTRATION NUMBER: 32,714
REFERENCE/DOCKET NUMBER: D-5350
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713/651-5325
TELEFAX: 713/651-5246
TELEX: 762829
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3765 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-07-751-891B-1

Query Match 4.0%; Score 20; DB 3; Length 3765;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 458 CTTTCTTTCTTTGTTT 477
DB 3159 CTTTCTTTCTTTGTTT 3178

RESULT 5

US-08-131-365B-37
Sequence 37, Application US/08131365B
Patent No. 5527690
GENERAL INFORMATION:
APPLICANT: Brown, Michael S.
APPLICANT: Briggs, Michael R.
APPLICANT: Wang, Xiaodong
APPLICANT: Goldstein, Joseph L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
TO STEROL REGULATORY ELEMENT BINDING

TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/131,365B
FILING DATE: 01-OCT-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UTSD:372/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 4154 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
FEATURE:
NAME/KEY: CDS
LOCATION: 167..3607
US-08-131-365B-37

Query Match 4.0%; Score 20; DB 1; Length 4154;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 337 AGCAGTGCCTGCGCCAGGCTG 356
Db 2393 AGCAGTGCCTGCGCCAGGCTG 2412

RESULT 6
US-08-668-123-37
Sequence 37, Application US/08668123
Patent No. 5891631
GENERAL INFORMATION:
APPLICANT: Brown, Michael S.
APPLICANT: Briggs, Michael R.
APPLICANT: Wang, Xiaodong
APPLICANT: Goldstein, Joseph L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING
NUMBER OF SEQUENCES: 64
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/668,123
FILING DATE: 14-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/131,365
FILING DATE: 01-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UTSD:372/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 4154 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
FEATURE:
NAME/KEY: CDS
LOCATION: 167..3607
US-08-668-123-37

Query Match 4.0%; Score 20; DB 2; Length 4154;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 337 AGCAGTGCCTGCGCCAGGCTG 356
Db 2393 AGCAGTGCCTGCGCCAGGCTG 2412

RESULT 7
US-08-455-073A-1
Sequence 1, Application US/08455073A
Patent No. 5876949
GENERAL INFORMATION:
APPLICANT: Gideon Dreyfuss
APPLICANT: Mikiko C. Siomi
APPLICANT: Yan Zhang
TITLE OF INVENTION: Fragile X Related Proteins, Compositions And Methods
TITLE OF INVENTION: Of Making And Using The Same
NUMBER OF SEQUENCES: 13
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5876949ris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,073A
FILING DATE: 31-MAY-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: UPN-2201
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4362 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 220..2118
US-08-455-073A-1

Query Match 4.0%; Score 20; DB 2; Length 4362;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 458 CTTTCTTTCTTTGTTTTT 477
DB 3303 CTTTCTTTCTTTGTTTTT 3322

RESULT 8

US-09-976-594-99
Sequence 99, Application US/09976594
Patent No. 6673549
GENERAL INFORMATION:
APPLICANT: Furness, Michael
APPLICANT: Buchbinder, Jenny
TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
FILE REFERENCE: PA-0041 US
CURRENT APPLICATION NUMBER: US/09/976,594
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 60/240,409
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 1143
SOFTWARE: PERL Program
SEQ ID NO 99
LENGTH: 6732
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6673549 272843.14
NAME/KEY: unsure
LOCATION: 444, 4904, 4907, 4910, 4914
OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-99

Query Match 4.0%; Score 20; DB 4; Length 6732;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 337 AGCAGTGCCTCCGAGGCTG 356
DB 2484 AGCAGTGCCTCCGAGGCTG 2503

RESULT 9

US-09-790-988-1
Sequence 1, Application US/09790988
Patent No. 6632935
GENERAL INFORMATION:
APPLICANT: SHIGENOBU, SHUJI
APPLICANT: WATANABE, HIDEKI
APPLICANT: HATTORI, MASAHIRA
APPLICANT: SAKAKI, YOSHIYUKI
TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
FILE REFERENCE: 081356/0159
CURRENT APPLICATION NUMBER: US/09/790,988
CURRENT FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: JP2000-107160
PRIOR FILING DATE: 2000-04-07
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 640681
TYPE: DNA

ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 4.0%; Score 20; DB 4; Length 640681;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 TTTTAAAAAAATATAA 491
DB 591945 TTTTAAAAAAATATAA 591964

RESULT 10

US-08-956-171E-3987
Sequence 3987, Application US/08956171E
Patent No. 6593114
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
APPLICANT: Gil H. Choi
APPLICANT: Patrick S. Dillon
APPLICANT: Craig A. Rosen
APPLICANT: Steven C. Barash
APPLICANT: Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 3987:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 3987:
US-08-956-171E-3987

Query Match 3.8%; Score 19; DB 4; Length 400;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 238 AAGCAGCGCTCTCCAG 256
DB 332 AAGCAGCGCTCTCCAG 350

RESULT 11

US-09-621-976-19147

```
; Sequence 19147, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 19147
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-621-976-19147
```

```
Query Match
Best Local Similarity 3.8%; Score 19; DB 4; Length 480;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 459 TTTTCTTTCTTTGTTTTT 477
Db 22 TTTTCTTTCTTTGTTTTT 40
```

```
RESULT 12
US-09-621-976-18831/c
; Sequence 18831, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18831
; LENGTH: 522
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-621-976-18831
```

```
Query Match
Best Local Similarity 3.8%; Score 19; DB 4; Length 522;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 468 TTTGTTTTTTAAAAAAA 486
Db 100 TTTGTTTTTTAAAAAAA 82
```

```
RESULT 13
US-09-976-594-498/c
; Sequence 498, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 498
; LENGTH: 543
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 223416.15
; US-09-976-594-498
```

```
Query Match
Best Local Similarity 3.8%; Score 19; DB 4; Length 543;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 459 TTTTCTTTCTTTGTTTTT 477
Db 351 TTTTCTTTCTTTGTTTTT 333
```

```
RESULT 14
US-09-220-132-151/c
; Sequence 151, Application US/09220132
; Patent No. 650607
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220,132
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 151
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(546)
; OTHER INFORMATION: n = A,T,C or G
; US-09-220-132-151
```

```
Query Match
Best Local Similarity 3.8%; Score 19; DB 4; Length 546;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 468 TTTGTTTTTTAAAAAAA 486
Db 25 TTTGTTTTTTAAAAAAA 7
```

```
RESULT 15
US-09-486-382B-7
; Sequence 7, Application US/09486382B
; Patent No. 6388174
; GENERAL INFORMATION:
; APPLICANT: Hokko Chemical Industry Co., Ltd.
; TITLE OF INVENTION: MINISTRY OF AGRICULTURE, FORESTRY AND FISHERIES
; FILE REFERENCE: 10647
; CURRENT APPLICATION NUMBER: US/09/486,382B
; CURRENT FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: JP 9-235049
; PRIOR FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
```

NAME/KEY: promoter
LOCATION: (1)..(1101)
US-09-486-382B-7

Query Match 3.88; Score 19; DB 4; Length 1101;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 467 TTTGTTTTTTAAAAAA 485
Db 51 TTTGTTTTTTAAAAAA 69

Search completed: February 24, 2004, 03:24:18
Job time : 47.2027 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 23:36:30 ; Search time 190.827 Seconds

(Without alignments)
9138.208 Million cell updates/sec

Title: US-09-936-456-3

Perfect score: 498
Sequence: 1 ctctagcgtgcgcgtctgcc.....aaaaaaataaagtcgacc 498

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2308684 seqs, 1750822206 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*

10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*

14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*

15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*

16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*

17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	425	85.3	1474	15	US-10-264-237-1097 Sequence 1097, Ap
2	416	83.5	1421	14	US-10-187-657-2 Sequence 2, Appl
3	407	81.7	1420	14	US-10-187-657-4 Sequence 4, Appl
4	392	78.7	2684	15	US-10-094-749-1195 Sequence 1195, Ap
5	96	19.3	346	14	US-10-187-657-3 Sequence 3, Appl
6	35	7.0	439	15	US-10-027-632-91970 Sequence 91970, A
7	23	4.6	408	12	US-10-424-599-78751 Sequence 78751, A
8	22	4.4	822	15	US-10-027-632-172855 Sequence 172855,
9	22	4.4	822	15	US-10-027-632-172856 Sequence 172856,
10	22	4.4	911	15	US-10-027-632-29944 Sequence 29944, A
11	22	4.4	2000	15	US-10-260-238-2487 Sequence 2487, Ap
12	22	4.4	3243	9	US-09-764-847-1176 Sequence 1176, Ap
13	22	4.4	3243	14	US-10-092-154-1176 Sequence 1176, Ap
14	21	4.2	325	9	US-09-960-352-7720 Sequence 7720, Ap
15	21	4.2	372	12	US-10-424-599-47139 Sequence 47139, A

C	16	21	4.2	555	14	US-10-029-386-7229	Sequence 7229, Ap
C	17	21	4.2	569	10	US-09-918-995-3019	Sequence 3019, Ap
C	18	21	4.2	623	10	US-09-871-161-207	Sequence 207, App
C	19	21	4.2	691	15	US-10-027-632-8433	Sequence 8433, Ap
C	20	21	4.2	691	15	US-10-027-632-8434	Sequence 8434, Ap
C	21	21	4.2	742	15	US-10-027-632-99671	Sequence 99671, A
C	22	21	4.2	742	15	US-10-027-632-99672	Sequence 99672, A
C	23	21	4.2	1441	14	US-10-171-581-69	Sequence 69, Appl
C	24	21	4.2	1473	14	US-10-247-671-7	Sequence 7, Appl
C	25	21	4.2	1484	9	US-09-925-300-248	Sequence 248, App
C	26	21	4.2	3730	9	US-09-880-107-1603	Sequence 1603, Ap
C	27	21	4.2	13084	14	US-10-311-455-2152	Sequence 2152, Ap
C	28	21	4.2	15788	14	US-10-017-161-1571	Sequence 1571, Ap
C	29	21	4.2	17703	16	US-10-257-166-33	Sequence 33, Appl
C	30	21	4.2	3673778	14	US-10-312-841-1	Sequence 1, Appl
C	31	20	4.0	307	12	US-10-424-599-54359	Sequence 54359, A
C	32	20	4.0	310	10	US-09-803-719-2001	Sequence 2001, Ap
C	33	20	4.0	331	12	US-10-424-599-133405	Sequence 133405,
C	34	20	4.0	385	12	US-10-424-599-129561	Sequence 129561,
C	35	20	4.0	397	10	US-09-918-995-36643	Sequence 36643, A
C	36	20	4.0	412	10	US-09-918-995-8664	Sequence 8664, Ap
C	37	20	4.0	415	15	US-10-027-632-24297	Sequence 24297, A
C	38	20	4.0	415	15	US-10-027-632-24298	Sequence 24298, A
C	39	20	4.0	416	9	US-09-960-352-4584	Sequence 4584, Ap
C	40	20	4.0	421	12	US-10-424-599-385	Sequence 385, App
C	41	20	4.0	471	15	US-10-027-632-107124	Sequence 107124,
C	42	20	4.0	478	10	US-09-918-995-11329	Sequence 11329, A
C	43	20	4.0	485	12	US-10-424-599-130822	Sequence 130822,
C	44	20	4.0	495	15	US-10-242-535A-28130	Sequence 28130, A
C	45	20	4.0	509	12	US-10-424-599-107007	Sequence 107007,

ALIGNMENTS

RESULT 1

US-10-264-237-1097

Sequence 1097, Application US/10264237

PUBLICATION NO. US20040009491A1

GENERAL INFORMATION:

APPLICANT: Birs et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PA131P1

CURRENT APPLICATION NUMBER: US/10/264,237

PRIOR FILING DATE: 2002-10-04

PRIOR APPLICATION NUMBER: PCT/US01/16450

PRIOR FILING DATE: 2001-05-18

PRIOR APPLICATION NUMBER: US 60/205,515

PRIOR FILING DATE: 2000-05-19

NUMBER OF SEQ ID NOS: 2876

SOFTWARE: PatentIn Ver. 3.1

SEQ ID NO 1097

LENGTH: 1474

TYPE: DNA

ORGANISM: Homo sapiens

US-10-264-237-1097

Query Match	85.3%	Score 425;	DB 15;	Length 1474;
Best Local Similarity	100.0%	Pred. No. 8.1e-204;		
Matches 425;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	62	GC	CGCCGCTTCTGGTCA	CGCTGGCAACGGCGCTCTGTGCTCTTCTCGAGGGCCG 121
DB	1024	GC	CGCCGCTTCTGGTCA	CGCTGGCAACGGCGCTCTGTGCTCTTCTCGAGGGCCG 1083
QY	122	TG	TGAGTCTCCAGTATG	TTCGCGCCAGCGCTCTTCGACCCCTTCTGACCAAGGCCCA 181
DB	1084	TG	TGAGTCTCCAGTATG	TTCGCGCCAGCGCTCTTCGACCCCTTCTGACCAAGGCCCA 1143
QY	182	AG	AGACTGACGACGAGAG	AGAGGGGCTCACCTTTATCCTCGGCGACCCACTGCACAAGC 241
DB	1144	AG	AGACTGACGACGAGAG	AGAGGGGCTCACCTTTATCCTCGGCGACCCACTGCACAAGC 1203

QY	242	AGGCCGCTCTCCAGACTTAAATGTATCACCCTAACCCTGTGAGGGGGACCCCAATCTGG	301
Db	1204	AGGCCGCTCTCCAGACTTAAATGTATCACCCTAACCCTGTGAGGGGGACCCCAATCTGG	1263
QY	302	ACTCCTTCCCCGCCCTTGGGACATCCAGAGCCGGGAAGCAGTGCCCGCAGGCTGGGCCA	361
Db	1264	ACTCCTTCCCCGCCCTTGGGACATCCAGAGCCGGGAAGCAGTGCCCGCAGGCTGGGCCA	1323
QY	362	GGAGAGCTCCAGAAAGGACATGAGCGCTGCTGGCGCGAGGCTCCGACATCCGAGGCA	421
Db	1324	GGAGAGCTCCAGAAAGGACATGAGCGCTGCTGGCGCGAGGCTCCGACATCCGAGGCA	1383
QY	422	CCAGGAAAGTCTCTGGGGCGATCTGTAAATAAAACCTTTTCTTTGTTTTTAAAA	481
Db	1384	CCAGGAAAGTCTCTGGGGCGATCTGTAAATAAAACCTTTTCTTTGTTTTTAAAA	1443
QY	482	AAAAA 486	
Db	1444	AAAAA 1448	

```

RESULT 2
US-10-187-657-2
; Sequence 2, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy R.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 1421
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030068311A1 490106CBI
US-10-187-657-2

```

Query Match	83.5%	Score 416	DB 14	Length 1421
Best Local Similarity	100.0%	Pred. No. 2.8e-199		
Matches 416	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	62	GGCGCGCCTTCTGGGTCACGCTGGCAACCGGCGCTCTGTGCTCTTCTCGAGAGGGCCG	121	
Db	1006	GGCGCGCCTTCTGGGTCACGCTGGCAACCGGCGCTGTGCTCTTCTCGAGAGGGCCG	1065	
QY	122	TGGTAGTCTCCAGTATGTTGGGCCAGCGCTCTTGACACCTTCTGACCAAAAGCGCA	181	
Db	1066	TGGTAGTCTCCAGTATGTTGGGCCAGCGCTCTTGACACCTTCTGACCAAAAGCGCA	1125	
QY	182	AGGACTGCAGCCAGAGAGAGAGGGGCTCACCCTTTATCCTCGGCAACCCACTGCACAAGC	241	
Db	1126	AGGACTGCAGCCAGAGAGAGAGGGGCTCACCCTTTATCCTCGGCAACCCACTGCACAAGC	1185	
QY	242	AGGCGGCTCTCCCACTTAATAATGATTCACCACTAACCTGTAGGGGGACCCATCTGG	301	
Db	1186	AGGCGGCTCTCCCACTTAATAATGATTCACCACTAACCTGTAGGGGGACCCATCTGG	1245	
QY	302	ACTCCTTCCCCGCTTTGGACATCGCAGCGCGGAAGCAGTGCCTCGCAGGCTTGGCCA	361	
Db	1246	ACTCCTTCCCCGCTTTGGACATCGCAGCGCGGAAGCAGTGCCTCGCAGGCTTGGCCA	1305	
QY	362	GGAGAGCTCCAGGAAGGCACTGAGCGCTGTGGCGGAGGCTTGGACATCCGCAAGCA	421	

Db 1306 GGAGGCTCCAGGAAAGGCACTGAGCGCTGCGCGAGGCCCTGGACATCCGCAGGCA 1365

QY 422 CCAGGAAAGTCTCCCTGGGGGATCTGTAATAAACCTTTTCTTTGTTTTT 477

Db 1366 CCAGGAAAGTCTCCCTGGGGGATCTGTAATAAACCTTTTCTTTGTTTTT 1421

```

RESULT 3
US-10-187-657-4
; Sequence 4, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 1420
; TYPE: DNA
; ORGANISM: Homo sapiens.
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030068311A1 3221661CA2
US-10-187-657-4

```

Query Match	81.7%;	Score 407;	DB 14;	Length 1420;
Best Local Similarity	100.0%;	Pred. No. 9.5e-195;		
Matches 407;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 62	GCGCCGCTTCTGGGTACGGCTGGCAACCGGCGTCTGTGCTCTTCTCTCGAGGGCCG			121
Db 1006	GCGCCGCTTCTGGGTACGGCTGGCAACCGGCGTCTGTGCTCTTCTCTCGAGGGCCG			1065
QY 122	TGGTAGTCTCCAGTATGTTTCGGCCAGCGCTCTTGGACACCTTCTGACCAAGCGCCA			181
Db 1066	TGGTAGTCTCCAGTATGTTTCGGCCAGCGCTCTTGGACACCTTCTGACCAAGCGCCA			1125
QY 182	AGAGCTGACAGCAGAGAGAGAGGGGGCTCACCTCTTATCTCTGGCGACCCACTGCACAGC			241
Db 1126	AGGACTGCAGCAGAGAGAGAGGGGGCTCACCTCTTATCTCTGGCGACCCACTGCACAGC			1185
QY 242	AGCGCGCTCTCCAGACTTAAATGTATCAACCACTAACCTGTAGGGGGACCAATCTGG			301
Db 1186	AGCGCGCTCTCCAGACTTAAATGTATCAACCACTAACCTGTAGGGGGACCAATCTGG			1245
QY 302	ACTCCTTCCCGCGCTTGGACATCGCAGCGCGGGAAGCAGTGCCTCCAGGCTTGGCCA			361
Db 1246	ACTCCTTCCCGCGCTTGGACATCGCAGCGCGGGAAGCAGTGCCTCCAGGCTTGGCCA			1305
QY 362	GGAGAGCTCGAGGAAGGGCAGCTGAGCGCTGTGCGGAGGCTTGGACATCCGAGGCA			421
Db 1306	GGAGAGCTCGAGGAAGGGCAGCTGAGCGCTGTGCGGAGGCTTGGACATCCGAGGCA			1365
QY 422	CCAGGAAAGTCTCTCTGGGGCGATCTGTAATAAACCTTTTCTT 468			
Db 1366	CCAGGAAAGTCTCTCTGGGGCGATCTGTAATAAACCTTTTCTT 1412			

RESULT 4
US-10-094-749-1195/C
; Sequence 1195, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:


```

; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1195
; LENGTH: 2684
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-1195
```

```

Query Match      78.7%; Score 392; DB 15; Length 2684;
Best Local Similarity 100.0%; Pred. No. 3.4e-187;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 91 GGGCTCCTGTGCTCTTCTCGAGAGGGCGGTGAGTCTCCAGTATGTTGGCCGAGC 150
DB 2343 GGGCTCCTGTGCTCTTCTCGAGAGGGCGGTGAGTCTCCAGTATGTTGGCCGAGC 2284
QY 151 GCTCTTCGACCCCTTCTGAGACCAAGCCGCAAGGACTGCAGCCAGAGAGAGGGGGCTCA 210
DB 2283 GCTCTTCGACCCCTTCTGAGACCAAGCCGCAAGGACTGCAGCCAGAGAGAGGGGGCTCA 2224
QY 211 CCTCTTATCTCGGCGAGCCCACTGCACAAGCAGGGCGCTCTCCAGACTTAAATGTATC 270
DB 2223 CCTCTTATCTCGGCGAGCCCACTGCACAAGCAGGGCGCTCTCCAGACTTAAATGTATC 2164
QY 271 ACCACTAAGCTGTGAGGGGAGACCAATCTGACTCTTCCCGCTTGGGACATCGCAGG 330
DB 2163 ACCACTAAGCTGTGAGGGGAGACCAATCTGACTCTTCCCGCTTGGGACATCGCAGG 2104
QY 331 CCGGAGAGAGTGGCCCGCAGGCTGGGCGAGAGAGTCCAGAGGGGCACTGAGCGCT 390
DB 2103 CCGGAGAGAGTGGCCCGCAGGCTGGGCGAGAGAGTCCAGAGGGGCACTGAGCGCT 2044
QY 391 GCTGGCGGAGGCTCGGACATCCGAGGACACGAGGAAGTCTCCTGGGGGCACTGTGTA 450
DB 2043 GCTGGCGGAGGCTCGGACATCCGAGGACACGAGGAAGTCTCCTGGGGGCACTGTGTA 1984
QY 451 AATAAACCTTTTCTTTGTTTTTAAAAA 482
DB 1983 AATAAACCTTTTCTTTGTTTTTAAAAA 1952
```

```

RESULT 5
US-10-187-657-3
; Sequence 3, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
```

```

; APPLICANT: Baughn, Mariah R.
; APPLICANT: AZIMZAI, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 3
; LENGTH: 346
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Indyte ID No. US20030068311A1 1752794F6
; NAME/KEY: unsure
; LOCATION: 91, 188, 206, 291, 337
; OTHER INFORMATION: a, t, c, g, or other
US-10-187-657-3
```

```

Query Match      19.3%; Score 96; DB 14; Length 346;
Best Local Similarity 100.0%; Pred. No. 4.3e-38;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 82 CTGGCAACCGGCTCTGCTCTCTCTCGAGAGGGCGGTGAGTCTCCAGTATGTT 141
DB 92 CTGGCAACCGGCTCTGCTCTCTCTCGAGAGGGCGGTGAGTCTCCAGTATGTT 151
QY 142 CGGCCAGCGCTCTCGACACCTTCTGACCAAGC 177
DB 152 CGGCCAGCGCTCTCGACACCTTCTGACCAAGC 187
```

```

RESULT 6
US-10-027-632-91970
; Sequence 91970, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 91970
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(439)
; OTHER INFORMATION: n = A, T, C or G
```

US-10-027-632-91970

Query Match
Best Local Similarity 7.0%; Score 35; DB 15; Length 439;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 CCGCTAGGCTCCTCCGCGCTCACCCTCAGTACG 60
DB 285 CCGCTAGGCTCCTCCGCGCTCACCCTCAGTACG 319

RESULT 7

US-10-424-599-78751/c
; Sequence 78751, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 78751
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(408)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_42127C.1
US-10-424-599-78751

Query Match
Best Local Similarity 4.6%; Score 23; DB 12; Length 408;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 TTGTTTTTAAAAAATAAAA 491
DB 86 TTGTTTTTAAAAAATAAAA 64

RESULT 8

US-10-027-632-172855/c
; Sequence 172855, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 172855
; LENGTH: 822
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-172855

Query Match
Best Local Similarity 4.4%; Score 22; DB 15; Length 822;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 TTTTTCCTTTGTTTTTAAA 480
DB 492 TTTTTCCTTTGTTTTTAAA 471

RESULT 9

US-10-027-632-172856/c
; Sequence 172856, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 172856
; LENGTH: 822
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-172856

Query Match
Best Local Similarity 4.4%; Score 22; DB 15; Length 822;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 TTTTTCCTTTGTTTTTAAA 480
DB 492 TTTTTCCTTTGTTTTTAAA 471

RESULT 10

US-10-027-632-29944
; Sequence 29944, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20

```
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29944
; LENGTH: 911
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-29944
```

```
Query Match          4.4%; Score 22; DB 15; Length 911;
Best Local Similarity 100.0%; Pred. No. 0.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 450 AAATTAACCTTTTCTTTG 471
Db 612 AAATTAACCTTTTCTTTG 633
```

```
RESULT 11
US-10-260-238-2487
; Sequence 2487, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Rieke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 2487
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-260-238-2487
```

```
Query Match          4.4%; Score 22; DB 15; Length 2000;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 472 TTTTAAAAAATAAAGT 493
Db 843 TTTTAAAAAATAAAGT 864
```

```
RESULT 12
US-09-764-847-1176/c
; Sequence 1176, Application US/09764847
; Patent No. US20020132767A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1176
; LENGTH: 3243
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1176
```

```
Query Match          4.4%; Score 22; DB 9; Length 3243;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 370 CCAGGAGGCGACTGAGCGCTG 391
Db 2584 CCAGGAGGCGACTGAGCGCTG 2563
```

```
RESULT 13
US-10-092-154-1176/c
; Sequence 1176, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1176
; LENGTH: 3243
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1176
```

```
Query Match          4.4%; Score 22; DB 14; Length 3243;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 370 CCAGGAGGCGACTGAGCGCTG 391
Db 2584 CCAGGAGGCGACTGAGCGCTG 2563
```

```
RESULT 14
US-09-960-352-7720/c
; Sequence 7720, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 7720
; LENGTH: 325
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 33-LIB34-009-Q1-E1-A2
US-09-960-352-7720
```

Query Match 4.2%; Score 21; DB 9; Length 325;
 Best Local Similarity 100.0%; Pred. No. 2.6;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 CCTTTTCTTTGTTTTT 477
 DB 320 CCTTTTCTTTGTTTTT 300

RESULT 15

US-10-424-599-47139/C
 ; Sequence 47139, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 47139
 ; LENGTH: 372
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_142572C.1
 US-10-424-599-47139

Query Match 4.2%; Score 21; DB 12; Length 372;
 Best Local Similarity 100.0%; Pred. No. 2.6;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 TTTTTTCTTTGTTTTTAA 479
 DB 353 TTTTTTCTTTGTTTTTAA 333

Search completed: February 24, 2004, 03:37:00
 Job time : 193.827 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 23:20:25 ; Search time 1266.96 Seconds

(without alignments)
11737.841 Million cell updates/sec

Title: US-09-936-456-3

Perfect score: 498
Sequence: 1 ccttagcgtgcgcgtcgcgc...aaaaaaataaagtcgacc 498

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hcc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
C	1	419	84.1	446	9	A1984141	A1984141 wu21c02.x
C	2	419	84.1	515	9	A1826629	A1826629 wk35e04.x
C	3	417	83.7	434	9	A1991272	A1991272 wu41h04.x
C	4	416	83.5	510	9	A1304327	A1304327 qo57h06.x

C	5	414	83.1	414	9	AA593860	nm19f03.s
C	6	408	81.9	537	9	A1821606	A1821606 nk08a11.x
C	7	405	81.3	405	9	A1274929	q149c11.x
C	8	388	77.9	538	9	A1660493	we67h02.x
C	9	382	76.7	411	9	A1983793	wu20c09.x
C	10	381	76.5	381	9	A1281211	qk58e08.x
C	11	375	75.3	573	14	CB854121	CB854121 UI-CF-DU1
C	12	374	75.1	405	9	A1346155	qp43f12.x
C	13	374	72.9	486	9	A1650560	we68b08.x
C	14	363	72.9	737	14	CB305399	UI-CF-EN1
C	15	348	69.9	737	14	BM987789	UI-H-CO0-
C	16	332	66.7	466	9	AA573825	nk08a11.s
C	17	308	61.8	324	12	BM987789	UI-H-CO0-
C	18	291	58.4	348	9	AM050605	wz19b11.x
C	19	280	56.2	690	12	BM977010	UI-CF-EN1
C	20	277	55.6	277	9	A1732165	ne17a03.x
C	21	270	54.2	687	14	CB850544	UI-CF-EN1
C	22	260	52.2	260	9	AM009962	w889h06.x
C	23	252	50.6	252	9	A1262416	qk38e04.x
C	24	185	37.1	539	9	A1924216	wn03b10.x
C	25	183	36.7	593	12	BM973444	UI-CF-EC1
C	26	166	33.3	315	9	AA618335	ng15g11.s
C	27	149	29.9	274	9	AA469031	ne17a03.s
C	28	127	25.5	505	9	A1791844	nk08a11.y
C	29	94	18.9	602	12	BQ017315	UI-H-DT1-
C	30	88	17.7	151	9	A1695625	we50h04.x
C	31	88	17.7	378	10	BF917041	IL3-DT011
C	32	44	8.8	119	9	A1921289	wo22h06.x
C	33	36	7.2	721	14	CA444325	UI-H-DT1-
C	34	25	5.0	670	28	BH130955	ENTNS27TR
C	35	25	5.0	778	12	BM393302	50071-2-8
C	36	25	5.0	778	12	BM395348	50072-2-8
C	37	25	5.0	866	28	BH136340	ENTM058TR
C	38	25	5.0	866	29	CNS02EYL	AL194358 Tetradon
C	39	25	5.0	871	28	AZ678595	ENT1220TR
C	40	25	5.0	898	28	BH157379	ENTTP92TR
C	41	25	5.0	929	28	AZ680299	ENTH009TR
C	42	25	5.0	936	28	BH153523	ENTS234TR
C	43	24	4.8	368	13	BY459696	BY459696
C	44	24	4.8	935	28	AZ682089	ENTJ033TF
C	45	24	4.8	1120	9	A1480624	vh52d01.x

ALIGNMENTS

RESULT 1
A1984141/c 446 bp mRNA linear EST 27-OCT-1999
LOCUS wu21c02.x1 Soares Dieckgraebe colon_NHCD Homo sapiens cDNA clone
DEFINITION IMAGE:2520674 3', mRNA sequence.

ACCESSION A1984141
VERSION A1984141.1 GI:5811360
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 446)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov

FEATURES
source 1..446
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from G1bco.
Location/Qualifiers

/clone="IMAGE:2520674"
/tissue type="colonic mucosa from 3 patients with Crohn's disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares_Dieckgraefe_colon_NHCD"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACCAATCTGAAGTGGAGCGCGCCGCTCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieck@im.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 84.1%; Score 419; DB 9; Length 446;
Best Local Similarity 100.0%; Pred. No. 4.6e-164;
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 GCGCCGCTTCTGGGTCAAGCTGCAACCGCGCTCTGCTCTCTCTCGAGGGGCGG 121
DB 419 GCGCCGCTTCTGGGTCAAGCTGCAACCGCGCTCTGCTCTCTCTCGAGGGGCGG 360
QY 122 TGGTGAAGTCTCAGTATGTTGGCCGCGCTCTTGCACCCCTTCTGACCAAGCGCA 181
DB 359 TGGTGAAGTCTCAGTATGTTGGCCGCGCTCTTGCACCCCTTCTGACCAAGCGCA 300
QY 182 AGGACTGCAGCAGAGAGAGAGGGGGCTCACTCTTATCTCTGCGGACCCACTGCACAAGC 241
DB 299 AGGACTGCAGCAGAGAGAGAGGGGGCTCACTCTTATCTCTGCGGACCCACTGCACAAGC 240
QY 242 AGGCGGCTCTCCAGACTTAAATGTATCAACCACTAAGCTGTGAGGGGAGCCCAATCTGG 301
DB 239 AGGCGGCTCTCCAGACTTAAATGTATCAACCACTAAGCTGTGAGGGGAGCCCAATCTGG 180
QY 302 ACTCCTTCCCGCCTTGGACATCGCAGCCGGAAGCAGTGCAGCCGAGGCTGGGCCA 361
DB 179 ACTCCTTCCCGCCTTGGACATCGCAGCCGGAAGCAGTGCAGCCGAGGCTGGGCCA 120
QY 362 GGAGAGCTCCAGAAAGGCACTGAGCGCTGCTGCGGAGGCGCTCGACATCCGAGGCA 421
DB 119 GGAGAGCTCCAGAAAGGCACTGAGCGCTGCTGCGGAGGCGCTCGACATCCGAGGCA 60
QY 422 CCAGGGAAGTCTCTGGGGCGATCTGTAAATAAACCTTTTCTTTGTTTTTAA 480
DB 59 CCAGGGAAGTCTCTGGGGCGATCTGTAAATAAACCTTTTCTTTGTTTTTAA 1

RESULT 2

AI826629/c 515 bp mRNA linear EST 21-DEC-1999
LOCUS wk35e04.x1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:2417406 3',
DEFINITION mRNA sequence.
ACCESSION AI826629
VERSION AI826629.1 GI:5447300
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 515)
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

FEATURES

source

Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert length: 910 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 467.
Location/Qualifiers
1. 515

ORIGIN

Query Match 84.1%; Score 419; DB 9; Length 515;
Best Local Similarity 100.0%; Pred. No. 4.3e-164;
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 GCGCCGCTTCTGGGTCAAGCTGCAACCGCGCTCTGCTCTCTCTCGAGGGGCGG 121
DB 419 GCGCCGCTTCTGGGTCAAGCTGCAACCGCGCTCTGCTCTCTCTCGAGGGGCGG 360
QY 122 TGGTGAAGTCTCAGTATGTTGGCCGCGCTCTTGCACCCCTTCTGACCAAGCGCA 181
DB 359 TGGTGAAGTCTCAGTATGTTGGCCGCGCTCTTGCACCCCTTCTGACCAAGCGCA 300
QY 182 AGGACTGCAGCAGAGAGAGAGGGGGCTCACTCTTATCTCTGCGGACCCACTGCACAAGC 241
DB 299 AGGACTGCAGCAGAGAGAGAGGGGGCTCACTCTTATCTCTGCGGACCCACTGCACAAGC 240
QY 242 AGGCGGCTCTCCAGACTTAAATGTATCAACCACTAAGCTGTGAGGGGAGCCCAATCTGG 301
DB 239 AGGCGGCTCTCCAGACTTAAATGTATCAACCACTAAGCTGTGAGGGGAGCCCAATCTGG 180
QY 302 ACTCCTTCCCGCCTTGGACATCGCAGCCGGAAGCAGTGCAGCCGAGGCTGGGCCA 361
DB 179 ACTCCTTCCCGCCTTGGACATCGCAGCCGGAAGCAGTGCAGCCGAGGCTGGGCCA 120
QY 362 GGAGAGCTCCAGAAAGGCACTGAGCGCTGCTGCGGAGGCGCTCGACATCCGAGGCA 421
DB 119 GGAGAGCTCCAGAAAGGCACTGAGCGCTGCTGCGGAGGCGCTCGACATCCGAGGCA 60
QY 422 CCAGGGAAGTCTCTGGGGCGATCTGTAAATAAACCTTTTCTTTGTTTTTAA 480
DB 59 CCAGGGAAGTCTCTGGGGCGATCTGTAAATAAACCTTTTCTTTGTTTTTAA 1

RESULT 3

AI991272/c 434 bp mRNA linear EST 09-MAR-2000
LOCUS wu41h04.x1 Soares_Dieckgraefe_colon_NHCD Homo sapiens cDNA clone
DEFINITION IMAGE:2522647 3', mRNA sequence.
ACCESSION AI991272
VERSION AI991272.1 GI:5838177
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 434)

REFERENCE

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.

COMMENT

Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 558 Std Error: 0.00
Seq primer: -40UP from Gibco.

FEATURES

source

1..434
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2522647"
/tissue_type="colonic mucosa from 3 patients with Crohn's
disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares Dieckgraebe colon NHCD"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Tissue samples
provided by Dr. Brian Dieckgraebe (Washington University,
dieck@im.wustl.edu); colonic mucosa represents a range of
disease involvement from moderate to severe Crohn's
disease; samples include both perforating (fistulas) and
non-perforating samples. Library constructed by Bento
Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 83.7%; Score 417; DB 9; Length 434;
Best Local Similarity 100.0%; Pred. No. 3.2e-163;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 GCGCGCCTTGGGTACGCTGGCAACGGCGCTGCTTCTTCTCGAGAGGGCCG 121
Db 417 GCGCGCCTTGGGTACGCTGGCAACGGCGCTGCTTCTTCTCGAGAGGGCCG 358
QY 122 TGGTAGTCTCCAGTATGTTGGGCCAGCGCTCTTCGACCTTTCGACCAAGCGCA 181
Db 357 TGGTAGTCTCCAGTATGTTGGGCCAGCGCTCTTCGACCTTTCGACCAAGCGCA 298
QY 182 AGGACTGCAGCCAGAGAGAGGGGGCTCACTCTTATCTCGGCAACCACTGCACAAGC 241
Db 297 AGGACTGCAGCCAGAGAGAGGGGGCTCACTCTTATCTCGGCAACCACTGCACAAGC 238
QY 242 AGGCGCTCTCCAGACTTAAATGTATCACTAACCCTGTGAGGGGGAGCCCAATCTGG 301
Db 237 AGGCGCTCTCCAGACTTAAATGTATCACTAACCCTGTGAGGGGGAGCCCAATCTGG 178
QY 302 ACTCTTCCCGCCTTGGAGCATCGAGGCCGGAGAGCAGTGCCTCGCAGGCTGGGCCA 361
Db 177 ACTCTTCCCGCCTTGGAGCATCGAGGCCGGAGAGCAGTGCCTCGCAGGCTGGGCCA 118
QY 362 GGAGAGCTCCAGAAAGGCACTGAGCGCTGCTGGCGGAGGCTCGGACATCCGACAGCA 421
Db 117 GGAGAGCTCCAGAAAGGCACTGAGCGCTGCTGGCGGAGGCTCGGACATCCGACAGCA 58
QY 422 CCAGGAAAGTCTCTGGGCGATCTGTAATAAACCTTTTCTTTGTTT 478
Db 57 CCAGGAAAGTCTCTGGGCGATCTGTAATAAACCTTTTCTTTGTTT 1

RESULT 4

AI304327/c 510 bp mRNA linear EST 01-FEB-1999
LOCUS
DEFINITION
q057h06.x1 NCI-CGAP C08 Homo sapiens cDNA clone IMAGE:1912667 3'
similar to contains TAR1.t3 TAR1 repetitive element ;, mRNA
sequence.

ACCESSION

AI304327
AI304327.1 GI:3988016

VERSION

EST.

KEYWORDS

Homo sapiens (human)

SOURCE

Homo sapiens

ORGANISM

Homo sapiens (human)

REFERENCE

1 (bases 1 to 510)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 1487 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 476.

FEATURES

source

1..510
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1912667"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI CGAP C08"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 83.5%; Score 416; DB 9; Length 510;
Best Local Similarity 100.0%; Pred. No. 7.7e-163;
Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 GCGCGCCTTGGGTACGCTGGCAACGGCGCTGCTTCTTCTCGAGAGGGCCG 121
Db 416 GCGCGCCTTGGGTACGCTGGCAACGGCGCTGCTTCTTCTCGAGAGGGCCG 357
QY 122 TGGTAGTCTCCAGTATGTTGGGCCAGCGCTCTTCGACCTTTCGACCAAGCGCA 181
Db 356 TGGTAGTCTCCAGTATGTTGGGCCAGCGCTCTTCGACCTTTCGACCAAGCGCA 297
QY 182 AGGACTGCAGCCAGAGAGAGGGGGCTCACTCTTATCTCGGCAACCACTGCACAAGC 241
Db 296 AGGACTGCAGCCAGAGAGAGGGGGCTCACTCTTATCTCGGCAACCACTGCACAAGC 237
QY 242 AGGCGCTCTCCAGACTTAAATGTATCACTAACCCTGTGAGGGGGAGCCCAATCTGG 301
Db 236 AGGCGCTCTCCAGACTTAAATGTATCACTAACCCTGTGAGGGGGAGCCCAATCTGG 177
QY 302 ACTCTTCCCGCCTTGGAGCATCGAGGCCGGAGAGCAGTGCCTCGCAGGCTGGGCCA 361
Db 176 ACTCTTCCCGCCTTGGAGCATCGAGGCCGGAGAGCAGTGCCTCGCAGGCTGGGCCA 117

QY 362 GGAGAGCTCCAGGAAGGCACTGAGCGCTGCTGGCGGAGGCGCTTCGACATCCGACAGCA 421
|||||
Db 116 GGAGAGCTCCAGGAAGGCACTGAGCGCTGCTGGCGGAGGCGCTTCGACATCCGACAGCA 57
QY 422 CCAGGAAAGTCTCCTGGGGCGATCTGTAAATAAACCCTTTTCTTTTGTGTTTTT 477
|||||
Db 56 CCAGGAAAGTCTCCTGGGGCGATCTGTAAATAAACCCTTTTCTTTTGTGTTTTT 1

RESULT 5
AA593860/c

LOCUS AA593860 414 bp mRNA linear EST 25-SEP-1997
DEFINITION nm19f03.s1 NCI_CGAP_Col2 Homo sapiens cDNA clone IMAGE:1084349 3',
mRNA sequence.

ACCESSION AA593860
VERSION AA593860
KEYWORDS AA593860.1 GI:2408538
SOURCE EST.
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Unpublished (1997)
Email: cgapbs-remail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.

REFERENCE
AUTHORS CDNA Library Preparation: Stratagene, Inc.
TITLE CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 1204 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 406.
Location/Qualifiers

FEATURES
source

1. .414
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1084349"
/sex="mixed"
/tissue_type="colon tumor"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="NCI_CGAP_Col2"
/note="Organ: colon; Vector: Bluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Pooled colon tumors. 5' adaptor sequence: 5'
GAATTGGGACGAG 3' 3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."

ORIGIN

Query Match. 83.1%; Score 414; DB 9; Length 414;
Best Local Similarity 100.0%; Pred. No. 5.8e-162; Indels 0; Gaps 0;
Matches 414; Conservative 0; Mismatches 0;

QY 63 CGCCGCTTCTGGGTACCGCTGGCAACCGGCGTCTCTGCTCTTCTCTCGAGGGGCGGT 122
|||||
Db 414 CGCCGCTTCTGGGTACCGCTGGCAACCGGCGTCTCTGCTCTTCTCTCGAGGGGCGGT 355
QY 123 GGTGAGTCTTCAAGTATGTTGGCCCAAGGCTCTTCGACACCTTCGGAACCAAGCCGCA 182
|||||
Db 354 GGTGAGTCTTCAAGTATGTTGGCCCAAGGCTCTTCGACACCTTCGGAACCAAGCCGCA 295
QY 183 GGAATGCAGCAGAGAGAGAGGGGGCTACCTCTTATCTCGGCGAACCCTGCAACAAGCA 242
|||||
Db 294 GGAATGCAGCAGAGAGAGAGGGGGCTACCTCTTATCTCGGCGAACCCTGCAACAAGCA 235

QY 243 GCGCGCTCTCCAGACTTAATAATGATATCAACACTAACCCTGTGAGGGGAGACCAATCTGGA 302
|||||
Db 234 GCGCGCTCTCCAGACTTAATAATGATATCAACACTAACCCTGTGAGGGGAGACCAATCTGGA 175
QY 303 CTCCTTCCCGGCTTGGGACATCGCAGGCGCGGAAGCAGTGCCCCGCAAGGCTTGAGCCAG 362
|||||
Db 174 CTCCTTCCCGGCTTGGGACATCGCAGGCGCGGAAGCAGTGCCCCGCAAGGCTTGAGCCAG 115
QY 363 GAGAGCTCCAGGAAGGGCACTGAGCGCTGCTGGCGGAGGCTTCGACATCCGACAGCAC 422
|||||
Db 114 GAGAGCTCCAGGAAGGGCACTGAGCGCTGCTGGCGGAGGCTTCGACATCCGACAGCAC 55
QY 423 CAGGGAAGTCTCCTGGGGCGATCTGTAAATAAACCCTTTTCTTTTGTGTTTTT 476
|||||
Db 54 CAGGGAAGTCTCCTGGGGCGATCTGTAAATAAACCCTTTTCTTTTGTGTTTTT 1

RESULT 6
LOCUS A1821606/c 537 bp mRNA linear EST 13-DEC-1999
DEFINITION nk08a11.x5 NCI_CGAP_Col2 Homo sapiens cDNA clone IMAGE:1012892 3',
similar to contains TAR1.t3 TAR1 MER22 repetitive element ;, mRNA
sequence.

ACCESSION A1821606
VERSION A1821606.1 GI:5440685
KEYWORDS EST.
SOURCE EST.
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL
COMMENT Unpublished (1997)
Other ESTs: nk08a11.y5
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: Stratagene, Inc., David B. Krizman,
Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: National Cancer Institute, Cancer Genome
Anatomy Project (CGAP), Tumor Gene Index.
This read has been verified (found to hit its original self in the
correct orientation)
Insert Length: 741 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 459.
Location/Qualifiers

FEATURES
source

1. .537
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1012892"
/tissue_type="tumor"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="NCI_CGAP_Col2"
/note="Organ: colon; Vector: Bluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Bulk colon villous adenoma. 5' adaptor sequence:
5' GAATTGGGACGAG 3' 3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

ORIGIN

Query Match 81.9%; Score 408; DB 9; Length 537;

Best Local Similarity 100.0%; Pred. No. 1.6e-159;
Matches 408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 62 GCGCCGCTCTTGGGTACAGCTGGCAACGGCGCTCCTGCTCTTCTCGAGGGGCGG 121
    |||
Db 417 GCGCCGCTCTTGGGTACAGCTGGCAACGGCGCTCCTGCTCTTCTCGAGGGGCGG 358
    |||
OY 122 TGGTAGTCTCCAGTATGTTCCGCCCGAGCGCTCTTCCGACCCCTTGGACCAAGCGCCA 181
    |||
Db 357 TGGTAGTCTCCAGTATGTTCCGCCCGAGCGCTCTTCCGACCCCTTGGACCAAGCGCCA 298
    |||
OY 182 AGGACTGCAGCCAGAGAGAGAGGGGGCTCACTCTTATCTCGGCGACCCACTGCACAAGC 241
    |||
Db 297 AGGACTGCAGCCAGAGAGAGAGGGGGCTCACTCTTATCTCGGCGACCCACTGCACAAGC 238
    |||
OY 242 AGGCCGCTCTCCAGACTTAATGTATCAACCTAACCCTGTGAGGGGGACCCAACTCTGG 301
    |||
Db 237 AGGCCGCTCTCCAGACTTAATGTATCAACCTAACCCTGTGAGGGGGACCCAACTCTGG 178
    |||
OY 302 ACTCCTTCCCGCGCTTGGGACATCGCAGGCGGGAGAGAGTGCCCGCCAGGCGCTGGGCCA 361
    |||
Db 177 ACTCCTTCCCGCGCTTGGGACATCGCAGGCGGGAGAGAGTGCCCGCCAGGCGCTGGGCCA 118
    |||
OY 362 GGAGAGCTCCAGGAGAGGGGCACTGAGCGCTGTGGCGGAGGCGCTCGACATCCGAGGCA 421
    |||
Db 117 GGAGAGCTCCAGGAGAGGGGCACTGAGCGCTGTGGCGGAGGCGCTCGACATCCGAGGCA 58
    |||
OY 422 CCAGGGAAGTCTCTGGGGCGATCTGTAAATTAACCTTTTCTTT 469
    |||
Db 57 CCAGGGAAGTCTCTGGGGCGATCTGTAAATTAACCTTTTCTTT 10
    |||
```

RESULT 7
AI274929/c 405 bp mRNA linear EST 29-JAN-1999
LOCUS q149c11.x1 NCI_CGAP_C08 Homo sapiens cDNA clone IMAGE:1875668 3'
DEFINITION similar to contains TAR1.t3 TAR1 repetitive element ;, mRNA
sequence.

ACCESSION AI274929
VERSION AI274929.1 GI:3897203
KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 405)
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

TISSUE Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1458 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 395.

FEATURES

Location/Qualifiers
1..405
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1875668"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_C08"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a

modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo. "

Query Match 81.3%; Score 405; DB 9; Length 405;
Best Local Similarity 100.0%; Pred. No. 3.3e-158;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 73 TGGTCAAGCTGGCAACGGCGTCTGTGCTCTTCTCGAGGGGCGGTGAGTCTC 132
    |||
Db 405 TGGTCAAGCTGGCAACGGCGTCTGTGCTCTTCTCGAGGGGCGGTGAGTCTC 346
    |||
OY 133 CAGTATGTTGGCCCGAGCGCTCTTGCACCCCTTGTGACCAAGCGCCAGACTGCAGC 192
    |||
Db 345 CAGTATGTTGGCCCGAGCGCTCTTGCACCCCTTGTGACCAAGCGCCAGACTGCAGC 286
    |||
OY 193 CAGAGAGAGGGGCTCACCCTCTTATCTCTGCGAGCCACTGCACAGAGAGCGCTCTC 252
    |||
Db 285 CAGAGAGAGGGGCTCACCCTCTTATCTCTGCGAGCCACTGCACAGAGAGCGCTCTC 226
    |||
OY 253 CCAGACTTAATGTATACCACTAACCCTGTGAGGGGGACCAATCTGACTCTCTCC 312
    |||
Db 225 CCAGACTTAATGTATACCACTAACCCTGTGAGGGGGACCAATCTGACTCTCTCC 166
    |||
OY 313 GCCTTGGGACATCGCAGGCGCGGAGAGCAGTGCCCGCCAGGCGCTGGGCCAGAGAGCTCCA 372
    |||
Db 165 GCCTTGGGACATCGCAGGCGCGGAGAGCAGTGCCCGCCAGGCGCTGGGCCAGAGAGCTCCA 106
    |||
OY 373 GGAAGGCACTGAGCGCTGCTGCGCGCGAGGCGCTCGACATCCGACGACCAAGGAAAGT 432
    |||
Db 105 GGAAGGCACTGAGCGCTGCTGCGCGCGAGGCGCTCGACATCCGACGACCAAGGAAAGT 46
    |||
OY 433 CTCCTGGGGCGGATCTGTAAATTAACCTTTTCTTTCTTTT 477
    |||
Db 45 CTCCTGGGGCGGATCTGTAAATTAACCTTTTCTTTCTTTT 1
    |||
```

RESULT 8
AI660493/c 538 bp mRNA linear EST 18-DEC-1999
LOCUS we67h02.x1 Soares Dieckgraefe_colon_NHCD Homo sapiens cDNA clone
DEFINITION IMAGE:2346195 3' similar to contains TAR1.t3 TAR1 repetitive
element ;, mRNA sequence.

ACCESSION AI660493

VERSION AI660493.1 GI:4764063
KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 538)
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

TISSUE This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 676 Std Error: 0.00
Seq primer: -40UP from Gibco

High quality sequence stop: 463.

FEATURES

Location/Qualifiers
1..538
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2346195"

/tissue_type="colonic mucosa from 3 patients with Crohn's disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares Dieckgraefe_colon_NHCD"
/note="Organ: colon; Vector: pT73D-Fac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACCAATCTGAAGTGGAGCGGCCGCTTTTCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieck@im.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 77.9%; Score 388; DB 9; Length 538;
Best Local Similarity 100.0%; Pred. No. 3.5e-151;
Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GGGCTCTGTCCTCTTCTCGAGGGCGGTGTGAGTCTCCAGTATGTTGGCCAGC 150
DB 388 GGGCTCTGTCCTCTTCTCGAGGGCGGTGTGAGTCTCCAGTATGTTGGCCAGC 329
QY 151 GCTCTTCGACCCCTTCTGACCAAAAGCCCAAGAGTGCAGCCAGAGAGAGGGGGCTCA 210
DB 328 GCTCTTCGACCCCTTCTGACCAAAAGCCCAAGAGTGCAGCCAGAGAGAGGGGGCTCA 269
QY 211 CCTCTTATCTTCGCGACCACTGCACAAAGAGCGGCTCTCCAGACTTAAATGTATC 270
DB 268 CCTCTTATCTTCGCGACCACTGCACAAAGAGCGGCTCTCCAGACTTAAATGTATC 209
QY 271 ACCACTAAGCTGTAGGGGAGCCCAATGTGACTCCTTCCCGCTTGGACATCGCAGG 330
DB 208 ACCACTAAGCTGTAGGGGAGCCCAATGTGACTCCTTCCCGCTTGGACATCGCAGG 149
QY 331 CCGGGAAGCAGTGCCTCCGACAGGCTGGGCGAGAGACTCCAGAGAGGGCAGTGAAGCGCT 390
DB 148 CCGGGAAGCAGTGCCTCCGACAGGCTGGGCGAGAGACTCCAGAGAGGGCAGTGAAGCGCT 89
QY 391 GCTGGCGAGAGGCTCGACATCCGACAGGACCAAGGAAAGTCTCCTGGGCGATCTGTA 450
DB 88 GCTGGCGAGAGGCTCGACATCCGACAGGACCAAGGAAAGTCTCCTGGGCGATCTGTA 29
QY 451 AATAAACCTTTTCTTTGTTTTTA 478
DB 28 AATAAACCTTTTCTTTGTTTTTA 1

RESULT 9
AI983793/c 411 bp mRNA linear EST 27-OCT-1999
LOCUS AI983793
DEFINITION wu20c09.x1 Soares Dieckgraefe_colon_NHCD Homo sapiens cDNA clone
IMAGE:2520592 3', mRNA sequence.
ACCESSION AI983793
VERSION AI983793.1 GI:5811012
KEYWORDS EST,
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapdb-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 400.
Location/Qualifiers

FEATURES
source

1. 411
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2520592"
/tissue_type="colonic mucosa from 3 patients with Crohn's disease"

/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares Dieckgraefe_colon_NHCD"
/note="Organ: colon; Vector: pT73D-Fac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACCAATCTGAAGTGGAGCGGCCGCTTTTCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieck@im.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 76.7%; Score 382; DB 9; Length 411;
Best Local Similarity 100.0%; Pred. No. 1.3e-148;
Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 CTGGGTACAGCTGGCAACCGCGCTCTGCTCTTCTTCGAGGGCGGTGTGAGTCT 131
DB 411 CTGGGTACAGCTGGCAACCGCGCTCTGCTCTTCTTCGAGGGCGGTGTGAGTCT 352
QY 132 CCAATATGTTGCGCCCGACGCTCTTCCGACCCCTTCTGACCAAAAGCCCAAGAGACTGACG 191
DB 351 CCAATATGTTGCGCCCGACGCTCTTCCGACCCCTTCTGACCAAAAGCCCAAGAGACTGACG 292
QY 192 CCAAGAGAGAGGGGCTCAGCTCTTATCTCTGCGGACCACTGCACAGCAGGCGGCTCT 251
DB 291 CCAAGAGAGAGGGGCTCAGCTCTTATCTCTGCGGACCACTGCACAGCAGGCGGCTCT 232
QY 252 CCGAGACTTAAATGTATACCACTAAGCTGTAGGGGAGCCCAATCTGAGACTCCTTCCC 311
DB 231 CCGAGACTTAAATGTATACCACTAAGCTGTAGGGGAGCCCAATCTGAGACTCCTTCCC 172
QY 312 CGCCTTGGGACATCGCAGGCGGGGAGAGCAGTGCCTCCGAGGCTGGGCGAGAGAGCTCC 371
DB 171 CGCCTTGGGACATCGCAGGCGGGGAGAGCAGTGCCTCCGAGGCTGGGCGAGAGAGCTCC 112
QY 372 AGGAAGGGCAGTGAAGCGCTGTGCGGCGAGGCTCGGACATCCGAGGACCAAGGAAAG 431
DB 111 AGGAAGGGCAGTGAAGCGCTGTGCGGCGAGGCTCGGACATCCGAGGACCAAGGAAAG 52
QY 432 TCTCCTGGGGCGATCTGTAAT 453
DB 51 TCTCCTGGGGCGATCTGTAAT 30

RESULT 10
AI281211/c 381 bp mRNA linear EST 28-JAN-1999
LOCUS AI281211
DEFINITION qk58608.x1 NCI-CCAP C08 Homo sapiens cDNA clone IMAGE:187190 3',
similar to contains TAR1.t3 TAR1 repetitive element ;, mRNA
sequence.
ACCESSION AI281211
VERSION AI281211.1 GI:3919444
KEYWORDS EST,
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
1 (bases 1 to 381)
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1422 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 340.
Location/Qualifiers
1. 381

FEATURES
source

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1873190"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI CGAP Co8"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 76.5%; Score 381; DB 9; Length 381;
Best Local Similarity 100.0%; Pred. No. 3.4e-148;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 97 CTGTGCTCTCTCTCGAGGGGCGTGGAGTCTCCAGTATGTTGGGCCAGCGCTCTT 156
Db 381 CTGTGCTCTCTCTCGAGGGGCGTGGAGTCTCCAGTATGTTGGGCCAGCGCTCTT 322
QY 157 CGCACCTTCTGAGCAAAAGCGCAAGAGCTGAGCCAGAGAGAGGGGGCTCACTCTT 216
Db 321 CGCACCTTCTGAGCAAAAGCGCAAGAGCTGAGCCAGAGAGAGGGGGCTCACTCTT 262
QY 217 ATCTCGGCGACCCACTGACCAAGCAGCGCTCTCCAGACTTAATGTATCACCACT 276
Db 261 ATCTCGGCGACCCACTGACCAAGCAGCGCTCTCCAGACTTAATGTATCACCACT 202
QY 277 AACCTGTAGGGGAGCACTCTGACTCTTCCCGCTTGGAGCATCGAGCGCGGA 336
Db 201 AACCTGTAGGGGAGCACTCTGACTCTTCCCGCTTGGAGCATCGAGCGCGGA 142
QY 337 AGCAGTGCCTGCGCAGGCTGCGCCAGAGAGCTCCAGAAAGGACATGAGCGCTGCGC 396
Db 141 AGCAGTGCCTGCGCAGGCTGCGCCAGAGAGCTCCAGAAAGGACATGAGCGCTGCGC 82
QY 397 GCGAGGCTCTGAGCATCCGAGAGGACACGAGAAAGTCTCTGGGGCATCTGTAATATA 456
Db 81 GCGAGGCTCTGAGCATCCGAGAGGACACGAGAAAGTCTCTGGGGCATCTGTAATATA 22
QY 457 CCTTTTCTTTTCTTTTCTTTT 477
Db 21 CCTTTTCTTTTCTTTTCTTTT 1

```

RESULT 11

CB854121/c
LOCUS
DEFINITION
UI-CF-DUI-aal-1-16-0-UI.s1 UI-CF-DUI Homo sapiens cDNA clone
UI-CF-DUI-aal-1-16-0-UI 3', mRNA sequence.
ACCESSION
CB854121
VERSION
CB854121.1 GI:30044498
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
1 (bases 1 to 573)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE
PUBMED
COMMENT
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 533-573, >Alu (matched complement)
Seq primer: M13 FORWARD
POLYA=No.

FEATURES
source

Location/Qualifiers
1. 573
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-DUI-aal-1-16-0-UI"
/tissue_type="Primary Lung Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-DUI"
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-DUI is a normalized cDNA library containing the
following tissue(s): Primary Lung Epithelial Cells The
library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT7T3-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is GGCTGTAGGC.
TAG_SEQ=None found"

ORIGIN

Query Match 75.3%; Score 375; DB 14; Length 573;
Best Local Similarity 100.0%; Pred. No. 9e-146;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 108 CCTCGAGGGGCGGTGAGTCTCCAGTATGTTCCGCGCGCTCTTGCACCCCTTCT 167
Db 482 CCTCGAGGGGCGGTGAGTCTCCAGTATGTTCCGCGCGCTCTTGCACCCCTTCT 423
QY 168 GGACCAAGCGCGCAAGACTGACGCCAGAGAGAGAGGGGCTCACTCTTATCTCGGCGA 227
Db 422 GGACCAAGCGCGCAAGACTGACGCCAGAGAGAGAGGGGCTCACTCTTATCTCGGCGA 363

```


[illegible]

RESULT 12	
AI346155/c	
LOCUS	
DEFINITION	
AI346155	405 bp
qp43fi2.x1 NCI CGAP C08 Homo sapiens cDNA clone IMAGE:1925807 3'	mRNA linear EST 02-FEB-1999

ACCESSION AI346155
VERSION AI346155.1 GI:4083361

SOURCE	Homo sapiens (human)
--------	----------------------

REFERENCE
AUTHORS
TITLE

Eukaryotes; Metazoa; Choriata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Carnivora; Homini; Homidae; Homo.
1 (bases 1 to 405)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute Cancer Genome Anatomy Project (CGAP).

JOURNAL
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
COMMENT

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
 Insert length: 1488 Std Error: 0.00
 Seq primer: -40UP from Gidco
 High quality sequence stop:375.

source	Location/Qualifiers
FEATURES	
source	1. .405

```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1925807"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
```

/note="Origin: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match	75.1%;	Score 374;	DB 9;	Length 405;
-------------	--------	------------	-------	-------------

Best Local Similarity 100.0%; Pred. No. 2.8e-145;
Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	73	TGGGTCACGCTGGGCAACCGGCGCTCTGTGCTCTTCTCCGAGGGGCGGTGTGATCTC	132
Db	405	TGGGTACGCTGGCAACCGGCGTCTGTGCTCTTCTCCGAGGGGCGGTGTGATCTC	346
QY	133	CAGTATGTTGGGCCAGCGCTTTTCGACCCCTTCTGACCAAAAGCCCAAGACTGCAGC	192
Db	345	CAGTATGTTGGGCCAGCGCTTTTCGACCCCTTCTGACCAAAAGCCCAAGACTGCAGC	286
QY	193	CAGGAGAGAGGGGCTCACCTCTTATCTCTCGGCGAACCCACTGCACAGCAGGCCGCTCTC	252
Db	285	CAGGAGAGAGGGGCTCACCTCTTATCTCTCGGCGAACCCACTGCACAGCAGGCCGCTCTC	226
QY	253	CCAGACTTAAATGTATCACCACTAACCTGTGAGGGGGAGCCCAATCTGSACTCCTTCCCC	312
Db	225	CCAGACTTAAATGTATCACCACTAACCTGTGAGGGGGAGCCCAATCTGSACTCCTTCCCC	166
QY	313	GCCTTGGACATCGCAGGCCGGGAAAGCAGTGCCTGCCAGGCTTGAGGAGAGCTCCA	372
Db	165	GCCTTGGACATCGCAGGCCGGGAAAGCAGTGCCTGCCAGGCTTGAGGAGAGCTCCA	106
QY	373	GGAAGGCACTGAGCGCTGCTGGCGCGAGGCGCTCGGACATCCGACAGGCCACGAGGAAAGT	432
Db	105	GGAAGGCACTGAGCGCTGCTGGCGCGAGGCGCTCGGACATCCGACAGGCCACGAGGAAAGT	46
QY	433	CTCCTGGGGCGATC 446	
Db	45	CTCCTGGGGCGATC 32	

RESULT	13
A1660560/c	
LOCUS	A1660560
DEFINITION	we68b08.x1 Soares Dieckgraefe colon NHCD Homo sapiens EST 18-DEC-1999
	486 bp mRNA linear
	EST 18-DEC-1999
	cdna clone

DEFINITION we68b08.x1 Soares Dieckgrafe_colon NHCD Homo sapiens cDNA clone IMAGE:2346231 3' similar to contains TARI.t3 TARI repetitive element /, mRNA sequence.

ACCESSION	AI660560
VERSION	AI660560.1
KEYWORDS	GI:4764130
SOURCE	EST.
	Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE
1 (bases 1 to 486)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 486)	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.

This clone is available royalty-free through LINL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1433 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 447.

FEATURES	Location/Qualifiers
source	1. .486

```

/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="IMAGE:2346231"
/tissue_type="colonic mucosa from 3 patients with Crohn's disease"

```

`/clone_lib="Soares Dieckgraefe_colon_NHCD"`
`/name="Organ: colon; Vector: pT7T3D-pac (Pharmacia) with a`
`modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st`
`strand cDNA was primed with a Not I - oligo(dT) primer [5'`
`TGTTTCCCATCTCAGAAGTGGAGCGCGCGCTTTTTTTTTTTTT 3'],`
`double-stranded cDNA was ligated to Eco RI adaptors`
`(Pharmacia), digested with Not I and cloned into the Not I`

and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieck@im.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo.

ORIGIN

Query Match 72.9%; Score 363; DB 9; Length 486;
Best Local Similarity 99.8%; Pred. No. 9.7e-141;
Matches 413; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 73 TGGGTACGCTGGCAACCGGCGTCTGTGCTTCTTCGAGGGGCGGTGAGTCTC 132
Db 420 TGGGTACGCTGGCAACCGGCGTCTGTGCTTCTTCGAGGGGCGGTGAGTCTC 361
QY 133 CAGTATGTTGGCCCGAGCGCTCTTCGACACCTTCTGACCAAGCGCCAGACTGCAGC 192
Db 360 CAGTATGTTGGCCCGAGCGCTCTTCGACACCTTTCGACCAAGCGCCAGACTGCAGC 301
QY 193 CAGGAGAGAGGGGCTCACCCTTTATCTTCGCGGACCCACTGCACAAAGAGGCGCTCTC 252
Db 300 CAGGAGAGAGGGGCTCACCCTTTATCTTCGCGGACCCACTGCACAAAGAGGCGCTCTC 241
QY 253 CCAGACTTAAATGTATCACCACTAACCTGTGAGGGGGAACCAATCTGACTCTTCCC 312
Db 240 CCAGACTTAAATGTATCACCACTAACCTGTGAGGGGGAACCAATCTGACTCTTCCC 181
QY 313 GCCTTGGACATCGCAGGCGCGGGAAGCAGTGCCTCCGAGGCTGGCCAGAGAGCTCCA 372
Db 180 GCCTTGGACATCGCAGGCGCGGGAAGCAGTGCCTCCGAGGCTGGCCAGAGAGCTCCA 121
QY 373 GGAAGGGCACTGAGCGCTGCTGGCGCGAGGCTCGACATCCGACAGGCAACGAGAAAGT 432
Db 120 GGAAGGGCACTGAGCGCTGCTGGCGCGAGGCTCGACATCCGACAGGCAACGAGAAAGT 61
QY 433 CTCTGGGGCGATCTGTAATAAACCTTTTCTTTTCTTTTAAAAAAA 486
Db 60 CTCTGGGGCGATCTGTAATAAACCTTTTCTTTTCTTTTAAAAAAA 7

RESULT 14

CB305399 737 bp mRNA linear EST 04-MAR-2003
LOCUS UI-CF-EN1-aed-m-17-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone

DEFINITION

UI-CF-EN1-aed-m-17-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
CB305399
CB305399.1 GI:28845910

ACCESSION

CB305399.1 GI:28845910

VERSION

EST.

KEYWORDS

Homo sapiens (human)

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 737)

AUTHORS

Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE

Normalization and subtraction: two approaches to facilitate gene

JOURNAL

Genome Res. 6 (9), 791-806 (1996)

MEDLINE

97044477

PUBMED

8889548

Contact: McCray, PB
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com) or from Open Biosystems (www.openbiosystems.com).
The following repetitive elements were found in this cDNA sequence: 29-138, >MIR#SINE/MIR (matched complement) 195-290, >ALU 223-320, >ALU
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES

source

Location/Qualifiers
1..737

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-CF-EN1-aed-m-17-0-UI"

/tissue_type="Primary Lung Cystic Fibrosis Epithelial Cells"

/dev_stage="Adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone_lib="UI-CF-EN1"

/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I;

UI-CF-EN1 is a normalized cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to

Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an

oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not

I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The

sequence tag for this library is CTGCTCAGGT.

TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS

6hr to LPS 24h

TAG_LIB=UI-CF-EN1

TAG_SEQ=CTGCTCAGGT"

ORIGIN

Query Match 69.9%; Score 348; DB 14; Length 737;
Best Local Similarity 100.0%; Pred. No. 1.4e-134;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GCGCTCTGTGCTCTTCTCGAGAGGGCGGTGTGAGTCTCCAGTATGTCGCCAGC 150
Db 353 GCGCTCTGTGCTCTTCTCGAGAGGGCGGTGTGAGTCTCCAGTATGTCGCCAGC 412
QY 151 GCTCTTGACACCTCTTGACCAAGCGCCAGAGACTGACGACGAGAGAGGGGCTCA 210
Db 413 GCTCTTGACACCTCTTGACCAAGCGCCAGAGACTGACGACGAGAGAGGGGCTCA 472
QY 211 CCTCTTATCTCGGCGAACCACCTGACCAAGAGCGGCTCTCCAGACTTAAATGTATC 270
Db 473 CCTCTTATCTCGGCGAACCACCTGACCAAGAGCGGCTCTCCAGACTTAAATGTATC 532
QY 271 ACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGAGACATGCAGG 330
Db 533 ACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTTCCCGCTTGAGACATGCAGG 592
QY 331 CCGGGAAGCAGTGCCCGCCAGGCGCTGGGCGAGAGAGCTCCAGGAAGGGCACTGAGCGCT 390
Db 593 CCGGGAAGCAGTGCCCGCCAGGCGCTGGGCGAGAGAGCTCCAGGAAGGGCACTGAGCGCT 652
QY 391 GCTGGCGGAGGCGCTGAGACATCCGAGAGGCAACGAGGAAAGTCTCTG 438
Db 653 GCTGGCGGAGGCGCTGAGACATCCGAGAGGCAACGAGGAAAGTCTCTG 700

RESULT 15

AA573825/c

LOCUS AA573825

DEFINITION nk08a11.s1 NCT_CGAP_Co2 Homo sapiens cDNA clone IMAGE:1012892 3',

mRNA sequence.

ACCESSION AA573825
 VERSION AA573825.1 GI:2348340
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 466)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.

CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert length: 741 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 456.
Location/Qualifiers

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1012892"
/tissue_type="tumor"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="NCI CGAP Co2"
/note="Organ: colon; Vector: Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Bulk colon villous adenoma. 5' adaptor sequence:
5' GAATTCGGCAGAG 3' 3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

```

Query Match	66.7 %	Score 332	DB 9	Length 466
Best Local Similarity	100.0 %	Pred. No. 8.3e-128		
Matches 332	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	60	GAGCGCCGCTTCTGGGTCA	CGCTGSCAACCGGCGCTCTGTGCTCTTCTCTGGAGGGG	119
Db	417	GAGCGCCGCTTCTGGGTCA	CGCTGSCAACCGGCGCTCTGTGCTCTTCTCTGGAGGGG	358
QY	120	CGTGTGAGTCTCCAGTATG	TTCGSCCGGCGCTTTCGACCCCTCTTGACCAAGCGC	179
Db	357	CGTGTGAGTCTCCAGTATG	TTCGSCCGGCGCTTTCGACCCCTCTTGACCAAGCGC	298
QY	180	CAAGCACTGCAGCCAGAGAG	AGAGGGGCGCTCACCTCTTATCTCTCGCGACCCACTGCACAA	239
Db	297	CAAGCACTGCAGCCAGAGAG	AGAGGGGCGCTCACCTCTTATCTCTCGCGACCCACTGCACAA	238
QY	240	GCAGCGCGCTCTCCAGACT	TAAATGTATCACCACTAACCCTGTGAGGGGAGCCCAATCT	299
Db	237	GCAGCGCGCTCTCCAGACT	TAAATGTATCACCACTAACCCTGTGAGGGGAGCCCAATCT	178
QY	300	GGACTCCTTCCCGCGCTTG	GGACATCGCAGGCCGGGAAGCAGTGCCTCCGACGGCTGGGC	359
Db	177	GGACTCCTTCCCGCGCTTG	GGACATCGCAGGCCGGGAAGCAGTGCCTCCGACGGCTGGGC	118
QY	360	CAGGAGAGCTCCAGGAAGG	GACACTGACCGCTG 391	
Db	117	CAGGAGAGCTCCAGGAAGG	GACACTGACCGCTG 86	

Search completed: February 24, 2004, 03:21:11
Job time : 1268.96 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 19:34:53 ; Search time 1965.03 Seconds

(without alignments)
10984.465 Million cell updates/sec

Title: US-09-936-456-3

Perfect score: 498

Sequence: 1 ctctagcgctgcgcgtctctgcc.....aaaaataaagtgcacc 498

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_cm:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	498	100.0	498	6	AX035348	AX035348 Sequence
2	465.4	93.5	1441	6	AX035346	AX035346 Sequence
3	463.8	93.1	1740	9	BSM806241	BSM806241 Homo sapi
4	398.4	80.0	2684	6	AX714511	AX714511 Sequence
5	398.4	80.0	2684	9	AX056896	AX056896 Homo sapi
6	398.4	80.0	156534	2	AC009700	AC009700 Homo sapi
7	398.4	80.0	166937	2	AC012255	AC012255 Homo sapi
8	398.4	80.0	171444	2	AC087790	AC087790 Homo sapi
9	398.4	80.0	181312	9	AC091117	AC091117 Homo sapi
10	217	43.6	217	9	HS183H12F	Z57552 H.sapiens C
11	199.8	40.1	218	9	HS183A12R	Z59954 H.sapiens C
12	166	33.3	1923	9	BC029819	BC029819 Homo sapi
13	160.6	32.2	1354	10	BC031111	BC031111 Mus muscu
14	117.8	23.7	229583	10	AL844566	AL844566 Mouse DNA
15	113.6	22.8	5851	4	AF547266	AF547266 Sus scrof
16	95	19.1	231868	2	AC118124	AC118124 Rattus no
17	50.2	10.1	125020	9	AF429315	AF429315 Homo sapi
18	48.2	9.7	125020	9	AF429315	AF429315 Homo sapi
19	46	9.2	1762	9	BC064138	BC064138 Homo sapi
20	46	9.2	3180	6	BD159535	BD159535 Sequence
21	46	9.2	3180	6	BD159535	BD159535 Primer fo
22	46	9.2	3180	9	AK022395	AK022395 Homo sapi
23	46	9.2	119347	9	AC020934	AC020934 Homo sapi
24	46	9.2	147750	9	AC018761	AC018761 Homo sapi
25	41.6	8.4	299425	1	AP005049	AP005049 Streptomy
26	40.2	8.1	33676	1	SCARD1GN	X84374 Saccharothr
27	40	8.0	3728	1	AB110077	AB110077 Streptomy
28	39.8	8.0	52359	2	AC010772	AC010772 Homo sapi
29	39.8	8.0	244105	2	AC098897	AC098897 Rattus no
30	39.6	8.0	160513	2	AC113563	AC113563 Caris fam
31	39.2	7.9	42338	9	HTMQ16F8	D85922 Homo sapien
32	39	7.8	117974	8	AC105730	AC105730 Oryza sat
33	39	7.8	166282	8	AC134233	AC134233 Oryza sat
34	39	7.8	182010	2	AC016767	AC016767 Homo sapi
35	38.8	7.8	2499	6	BD180276	BD180276 Highly th
36	38.8	7.8	34721	9	HS315G5	AL031708 Human DNA
37	38.8	7.8	270150	9	AE006639	AE006639 Homo sapi
38	38.6	7.8	2000	6	AX655393	AX655393 Sequence
39	38.6	7.8	166889	2	AC129938	AC129938 Mus muscu
40	38.6	7.8	218508	10	AC113497	AC113497 Mus muscu
41	38.6	7.8	299300	1	AP005026	AP005026 Streptomy
42	38.2	7.7	1094	4	SSC440724	AJ440724 Sus scrof
43	38.2	7.7	187080	9	AP000977	AP000977 Homo sapi
44	38.2	7.7	187917	2	AC025321	AC025321 Homo sapi
45	38.2	7.7	201509	2	AC145356	AC145356 Gorilla g

ALIGNMENTS

RESULT 1
AX035348
LOCUS AX035348 498 bp DNA linear PAT 15-NOV-2000
DEFINITION Sequence 3 from Patent WO0053748.
ACCESSION AX035348
VERSION AX035348.1 GI:11191065
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Bruck,C.E., Vinals,Y.D., Coche,T. and Casseart,J.P.
TITLE Novel compounds
JOURNAL Patent: WO 0053748-A 3 14-SEP-2000;

BRUCK CLAUDINE ELVIRE MARIE (BE) ; SMITHKLINE BEECHAM BIOLOG (BE) ;
VINALS Y DE BASSOLS CARLOTA (BE) ; COCHE THIERRY (BE) ; CASSART
JEAN POL (BE)

FEATURES

Location/Qualifiers
1..498
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match

100.0%; Score 498; DB 6; Length 498;

Best Local Similarity 100.0%; Pred. No. 9.6e-113;

Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CTCTAGCGGTGCGGCTCTGCGGCTCCGCTAGGCTCTCCGCGCTCACCCTCAGTACG 60
    |||
Db 1 CTCTAGCGGTGCGGCTCTGCGGCTCCGCTAGGCTCTCCGCGCTCACCCTCAGTACG 60
QY 61 AGCGCCGCTTCTGGGTCAAGCTGGCAACCGCGCTGTGCTCTTCTCTGGAGGGGCC 120
    |||
Db 61 AGCGCCGCTTCTGGGTCAAGCTGGCAACCGCGCTGTGCTCTTCTCTGGAGGGGCC 120
QY 121 GTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTCCACCTCTTGACCAAAAGCGCC 180
    |||
Db 121 GTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTCCACCTCTTGACCAAAAGCGCC 180
QY 181 AAGGACTGCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
    |||
Db 181 AAGGACTGCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY 241 CAGGCCGCTCTCCAGACTTAAATGTATCAACCACTAACCCTGTGAGGGGGACCAATCTG 300
    |||
Db 241 CAGGCCGCTCTCCAGACTTAAATGTATCAACCACTAACCCTGTGAGGGGGACCAATCTG 300
QY 301 GACTCTCTTCCCGCTTGGGACATGCGAGGCGCGGAGAGCAGTGGCCCGCAGGCTTGCGCC 360
    |||
Db 301 GACTCTCTTCCCGCTTGGGACATGCGAGGCGCGGAGAGCAGTGGCCCGCAGGCTTGCGCC 360
QY 361 AGGAGAGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
    |||
Db 361 AGGAGAGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY 421 ACCAGGGAAGTCTCTGAGGGCGGATCTGTAATTAACCTTTTCTTTTCTTTTAA 480
    |||
Db 421 ACCAGGGAAGTCTCTGAGGGCGGATCTGTAATTAACCTTTTCTTTTCTTTTAA 480
QY 481 AAAAAATAAAGTCGACC 498
    |||
Db 481 AAAAAATAAAGTCGACC 498
```

RESULT 2

AX035346 1441 bp DNA linear PAT 15-NOV-2000

LOCUS AX035346 Sequence 1 from Patent WO0053748.

DEFINITION

AX035346

ACCESSION

AX035346.1

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

BRUCK, C.E., Vinals, Y.D., Coche, T. and Cassart, J.P.

Novel compounds

Patent: WO 0053748-A 1 14-SEP-2000;

BRUCK CLAUDINE ELVIRE MARIE (BE) ; SMITHKLINE BEECHAM BIOLOG (BE) ;

VINALS Y DE BASSOLS CARLOTA (BE) ; COCHE THIERRY (BE) ; CASSART

JEAN POL (BE)

Location/Qualifiers

1..1441

/organism="Homo sapiens"

/mol_type="unassigned DNA"

ORIGIN

/db_xref="taxon:9606"

Query Match

93.5%; Score 465.4; DB 6; Length 1441;

Best Local Similarity 99.4%; Pred. No. 1.2e-104;

Matches 488; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

```
QY 1 CTCTAGCGGTGCGGCTCTGCGGCTCCGCTAGGCTCTCCGCGCTCACCCTCAGTACG 60
    |||
Db 939 CTCTAGCGGTGCGGCTCTGCGGCTCCGCTAGGCTCTCCGCGCTCACCCTCAGTACG 997
QY 61 AGCGCCGCTTCTGGGTCAAGCTGGCAACCGCGCTGTGCTCTTCTCTGGAGGGGCC 120
    |||
Db 998 -GCGCCGCTTCTGGGTCAAGCTGGCAACCGCGCTGTGCTCTTCTCTGGAGGGGCC 1056
QY 121 GTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTCCACCTCTTGACCAAAAGCGCC 180
    |||
Db 1057 GTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTCCACCTCTTGACCAAAAGCGCC 1116
QY 181 AAGGACTGCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
    |||
Db 1117 AAGGACTGCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1176
QY 241 CAGGCCGCTCTCCAGACTTAAATGTATCAACCACTAACCCTGTGAGGGGGACCAATCTG 300
    |||
Db 1177 CAGGCCGCTCTCCAGACTTAAATGTATCAACCACTAACCCTGTGAGGGGGACCAATCTG 1236
QY 301 GACTCTCTTCCCGCTTGGGACATGCGAGGCGCGGAGAGCAGTGGCCCGCAGGCTTGCGCC 360
    |||
Db 1237 GACTCTCTTCCCGCTTGGGACATGCGAGGCGCGGAGAGCAGTGGCCCGCAGGCTTGCGCC 1296
QY 361 AGGAGAGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
    |||
Db 1297 AGGAGAGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1356
QY 421 ACCAGGGAAGTCTCTGAGGGCGGATCTGTAATTAACCTTTTCTTTTCTTTTAA 480
    |||
Db 1357 ACCAGGGAAGTCTCTGAGGGCGGATCTGTAATTAACCTTTTCTTTTCTTTTAA 1416
QY 481 AAAAAATAAAGTCGACC 491
    |||
Db 1417 AAAAAATAAAGTCGACC 1427
```

RESULT 3

HSMB06241 1740 bp mRNA linear PRI 17-JUN-2003

LOCUS HSMB06241

DEFINITION

EX537581

ACCESSION

EX537581.1

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (17-JUN-2003) MIPS, Ingolstaedter Landstr.1, D-85764

Neuberger, GERMANY

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by GBF (National Research Centre for Biotechnology Ltd.,

Braunschweig/Germany) within the CDNA sequencing consortium of the

German Genome Project.

This clone (DKFZp686C04213) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further

information about the clone and the sequencing project is available

at http://mips.gsf.de/proj/cdna/.

Location/Qualifiers

1..1740

/organism="Homo sapiens"

/mol_type="unassigned DNA"

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="15q15"
/clone="DKFZp686C04213"
/tissue_type="human rectum tumor"
/clone_id="1686 (synonym: h1cc3). Vector pSport1_sfi; host
DH10B; sites SfiI + SfiIb"
/dev_stage="adult"
1694_ .1699
polyA_signal
polyA_site 1720
ORIGIN

```

```

Query Match      93.1%; Score 463.8; DB 9; Length 1740;
Best Local Similarity 99.2%; Pred. No. 3e-104;
Matches 487; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 CTCTAGCGTGCCTCTGCGCGCTCCCGCTAGGCTCCTCCGCGTCAACCACTAGTACG 60
    |||
DB 1246 CTCTAGCGTGCCTCTGCGCGCT-CCGCTAGGCTCCTCCGCGTCAACCACTAGTACG 1304

QY 61 AGCGCGCTTCTGGGTCAAGTGGCAACCGGCGTCTGTGCTCTTCTCGAGGGGCC 120
    |||
DB 1305 -GCGCGCTTCTGGGTCAAGTGGCAACCGGCGTCTGTGCTCTTCTCGAGGGGCC 1363

QY 121 GTGTGAGTCTCAGTATGTTCCGCCGCGCTCTTGCACCCCTTGACCAACCAAGCGCC 180
    |||
DB 1364 GTGTGAGTCTCAGTATGTTCCGCCGCGCTCTTGCACCCCTTGACCAACCAAGCGCC 1423

QY 181 AAGGACTGCAGCCAGAGAGAGGGGGCTCACTCTTATCTCTCGGCAACCACTGCAAG 240
    |||
DB 1424 AAGGACTGCAGCCAGAGAGAGGGGGCTCACTCTTATCTCTCGGCAACCACTGCAAG 1483

QY 241 CAGGCGCTCTCCAGACTTAAATGATACCACTAACCTGTGAGGGGGAACCAATCTG 300
    |||
DB 1484 CAGGCGCTCTCCAGACTTAAATGATACCACTAACCTGTGAGGGGGAACCAATCTG 1543

QY 301 GACTCCTTCCCGCGCTTGGGACATGCGAGCGCGGGAAGCAGTGCCTCGGCGCTGGGCC 360
    |||
DB 1544 GACTCCTTCCCGCGCTTGGGACATGCGAGCGCGGGAAGCAGTGCCTCGGCGCTGGGCC 1603

QY 361 AGGAGAGCTCCAGGAAGGGCACTGAGCGCTGCTGGCGGAGGCGCTCGGACATCGCAGGC 420
    |||
DB 1604 AGGAGAGCTCCAGGAAGGGCACTGAGCGCTGCTGGCGGAGGCGCTCGGACATCGCAGGC 1663

QY 421 ACCAGGGAAGTCTCTGGGGCGATCTGTAAATAAACCTTTTCTTTGTTTTTAA 480
    |||
DB 1664 ACCAGGGAAGTCTCTGGGGCGATCTGTAAATAAACCTTTTCTTTGTTTTTAA 1723

QY 481 AAAAAATATAA 491
    |||
DB 1724 AAAAAATATAA 1734

```

```

RESULT 4
AX714511/c 2684 bp DNA linear PAT 15-APR-2003
LOCUS AX714511
DEFINITION Sequence 1195 from Patent EP1293569.
ACCESSION AX714511
VERSION AX714511.1 GI:29889464
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
            Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
            Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
            Masuho,Y.
TITLE Full-length cDNAs
JOURNAL Patent: EP 1293569-A 1195 19-MAR-2003;
            Helix Research Institute (JP) ; Research Association for

```

```

Biotechnology (JP)
FEATURES
    Location/Qualifiers
        source
            1..2684
                /organism="Homo sapiens"
                /mol_type="unassigned DNA"
                /db_xref="taxon:9606"
ORIGIN

```

```

Query Match      80.0%; Score 398.4; DB 6; Length 2684;
Best Local Similarity 95.1%; Pred. No. 5e-88;
Matches 411; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 51 ACTCAGTACGAGCGCGCTTCTTGGGTACGCTGGCAACCGCGTCTGTGCTTCTCT 110
    |||
DB 2383 ACTCAGTACGAGCGCGCTTCTTGGGTACGCTGGCAACCGCGTCTGTGCTTCTCT 2324

QY 111 CGAGGGGCGGTGTGAGTCTTCCAGTATGTTGGCCAGCGCTTTGCACTTCTGGA 170
    |||
DB 2323 CGAGGGGCGGTGTGAGTCTTCCAGTATGTTGGCCAGCGCTTTGCACTTCTGGA 2264

QY 171 CCAAGCGCCAAAGACTGCAAGCCAGAGAGAGGGGGCTCACTTATCTCTCGGACCC 230
    |||
DB 2263 CCAAGCGCCAAAGACTGCAAGCCAGAGAGAGGGGGCTCACTTATCTCTCGGACCC 2204

QY 231 ACTGACAAAGCAGCGCGCTCTCCAGACTTAAATGATACCACTAAGCTGAGGGGG 290
    |||
DB 2203 ACTGACAAAGCAGCGCGCTCTCCAGACTTAAATGATACCACTAAGCTGAGGGGG 2144

QY 291 ACCCAATCTGAGTCTTCTCCCGCTTGGGACATGCGAGCGCGGGAAGCAGTGCCTGCA 350
    |||
DB 2143 ACCCAATCTGAGTCTTCTCCCGCTTGGGACATGCGAGCGCGGGAAGCAGTGCCTGCA 2084

QY 351 GGCTGGGCGCAGAGAGAGCTCCAGGAAGGGCACTGAGCGCTCTGCGCGAGCGCTCGGAC 410
    |||
DB 2083 GGCTGGGCGCAGAGAGAGCTCCAGGAAGGGCACTGAGCGCTCTGCGCGAGCGCTCGGAC 2024

QY 411 ATCCGACAGCACCAGGAAGTCTCTGGGGCGATCTGTAAATAAACCTTTTCTTTT 470
    |||
DB 2023 ATCCGACAGCACCAGGAAGTCTCTGGGGCGATCTGTAAATAAACCTTTTCTTTT 1964

QY 471 GTTTTTTAAAA 482
    |||
DB 1963 GTTTTTTAAAA 1952

```

```

RESULT 5
AK056896/c 2684 bp mRNA linear PRI 01-AUG-2002
LOCUS AK056896
DEFINITION Homo sapiens cDNA FLJ23334 f1s, clone PROST2005426.
ACCESSION AK056896
VERSION AK056896.1 GI:16552419
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,
            Ono,Y., Hotuta,T., Hiraoka,S., Murakawa,K., Takiguchi,S.,
            Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,
            Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,
            Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y.,
            Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y.,
            Sekine,M., Kikuchi,H., Kanda,K., Wagatsuna,M., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2684)
AUTHORS Isogai,T., Otsuki,T. and Sugiyama,T.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

```

COMMENT

(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3386)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: RAB and
HRI.

FEATURES

Source

```

Location/Qualifiers
1..2684
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PROST2005426"
/tissue_type="prostate"
/clone_lib="PROST2"
/note="cloning vector: pME18SFL3"
56..952

```

CDS

```

/note="unnamed protein product"
/codon_start=1
/protein_id="BAB71304.1"
/db_xref="GI:16552420"
/translation="MATLTGTFEPFYAGQKPTFPMDTTIASIMIFLTALATFIVILVGG
IRGTRFLFWLRLRYVTSLSFGAIIIGTPVQQLNETINNEBFTWRLGENVAEEYAKLE
KGLDDPVLYLAKEFTPRSPCGIYRQRLAGYTSAMLVAFLCWLINAVMLSPVLVY
GGWLLATGIFOLLALFFSMATSLTSPCDLGLGASVLRTHGPAFTLTGLICVIL
LGLAMVAHRMQRRLKAFPSVDSDECPMLEMSPEEGGLISPRYSMADSPKSQDLPD
SEASTKAYCKEAKPKDPDCAI"

```

ORIGIN

Query Match	80.0%;	Score 398.4;	DB 9;	Length 2684;
Best Local Similarity	95.1%;	Pred. No. 5e-88;		
Matches 411; Conservative	0;	Mismatches 21;	Indels 0;	Gaps 0;

QY	51	ACTCAGTACGAGCGCCGCTTCTGGATCAGCGCTGGCAACCGAGCTCCTGCTTTCT	110
Db	2383	ACTCCACATGCCCTCCTTTCTTTTCATCCCCACCGCCACAGGCGTCTGTGCTTTCT	23242
QY	111	CGAGAGGCGCGTGTGTAGTCTCCAGTATGTTGGGCCAGCGCTCTTCGACCCCTTCTGGA	170
Db	2323	CGAGAGGCGCGTGTGTAGTCTCCAGTATGTTGGGCCAGCGCTCTTCGACCCCTTCTGGA	2264
QY	171	CCAAAGCGCCAAAGSACTGCAGCCAGAGAGAGAGAGGGGGCTCACTCTTATCTCGCGCACCC	230
Db	2263	CCAAAGCGCCAAAGSACTGCAGCCAGAGAGAGAGGGGGCTCACTCTTATCTCGCGCACCC	2204
QY	231	ACTGCACAGCAGGCGCGTCTCCAGACTTAATAATGTATCAACCACTAACCTGTGAGGGG	290
Db	2203	ACTGCACAGCAGGCGCGTCTCCAGACTTAATAATGTATCAACCACTAACCTGTGAGGGG	21444
QY	291	ACCCAATCTGGACTCCTTCCCCCGCTTGGACATCGCAGGCGGGGAAGCAGTGCCCGCA	350
Db	2143	ACCCAATCTGGACTCCTTCCCCCGCTTGGACATCGCAGGCGGGGAAGCAGTGCCCGCA	2084
QY	351	GGCCTGGGCCAGGAGAGCTCCAGAGGAAGGCATGAGCGTGTGGCGCGAGGCTCGAC	410
Db	2083	GGCCTGGGCCAGGAGAGCTCCAGAGGAAGGCATGAGCGTGTGGCGCGAGGCTCGAC	2024
QY	411	ATCCGACAGCACGAGGAAAGTCTCCTGGGGCGATCTGTAAATAAACCTTTTTCTTT	470
Db	2023	ATCCGACAGCACGAGGAAAGTCTCCTGGGGCGATCTGTAAATAAACCTTTTTCTTT	1964
QY	471	GTTTTTTAAAA 482	
Db	1963	GTTTTTTAAAA 1952	

RESULT 6	AC009700	LOCUS	DEFINITION
	AC009700	156534 bp	DNA linear
		Homo sapiens chromosome 15 clone Rpl1-163P10 map 15, WORKING DRAFT SEQUENCE, 15 unordered pieces.	HTG 20-APR-2000

ACCESSION
VERSION
KEYWORDS
SOURCE

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

```

AC009700
AC009700.4 GI:7622346
HTG; HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens (human)
Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
1 (bases 1 to 156534)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-163p10
Unpublished
2 (bases 1 to 156534)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,N.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerry,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArrellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczyk,J., Liu,C., Locke,K., Macdonald,P.,
Marguis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Nilotf,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tefaye,S., Torturella-Willer,I., Vassilev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 20, 2000 this sequence version replaced gi:6056272.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L2336
Center clone name: 163_P10
----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 108837 bases at least Q40
Consensus quality: 130230 bases at least Q30
Consensus quality: 145227 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 155134; sum-of-contigs
Quality coverage: 3.2 in Q20 bases; agarose-fp
Quality coverage: 3.3 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1169: contig of 1169 bp in length
* 1170 1269: gap of 100 bp
* 1270 3335: contig of 2066 bp in length
* 3336 3435: gap of 100 bp
* 3436 5698: contig of 2263 bp in length
* 5699 5798: gap of 100 bp
* 5799 9216: contig of 3418 bp in length
* 9217 9316: gap of 100 bp
* 9317 15633: contig of 6317 bp in length
* 15634 15746: contig of 100 bp
* 15747 6013: gap of 6013 bp in length

```



```

* 21747 21846: gap of 100 bp
* 21847 30351: contig of 8505 bp in length
* 30352 30451: gap of 100 bp
* 30452 36509: contig of 6058 bp in length
* 36510 36609: gap of 100 bp
* 36610 45279: contig of 8670 bp in length
* 45280 45379: gap of 100 bp
* 45380 56851: contig of 11472 bp in length
* 56852 56952: gap of 100 bp
* 56952 67374: contig of 10423 bp in length
* 67375 67474: gap of 100 bp
* 67475 79534: contig of 12060 bp in length
* 79535 79635: gap of 100 bp
* 79635 95432: contig of 15798 bp in length
* 95433 95532: gap of 100 bp
* 95533 120801: contig of 25269 bp in length
* 120802 120901: gap of 100 bp
* 120902 156534: contig of 35633 bp in length.

FEATURES
    source
        1. 156534
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="15"
            /map="15"
            /clone="RP11-163P10"
            /clone_lib="RPC1-11 Human Male BAC"
            1. 1169
                /note="assembly_fragment"
                1270. 3335
                    /note="assembly_fragment"
                    3436. 5698
                        /note="assembly_fragment"
                        5799. 9216
                            /note="assembly_fragment"
                            9317. 15633
                                /note="assembly_fragment"
                                15734. 21746
                                    /note="assembly_fragment"
                                    21847. 30351
                                        /note="assembly_fragment"
                                        30452. 36509
                                            /note="assembly_fragment"
                                            clone_end:T7
                                            vector_side:left"
                                            36610. 45279
                                                /note="assembly_fragment"
                                                45380. 56851
                                                    /note="assembly_fragment"
                                                    56952. 67374
                                                        /note="assembly_fragment"
                                                        67475. 79534
                                                            /note="assembly_fragment"
                                                            79635. 95432
                                                                /note="assembly_fragment"
                                                                95533. 120801
                                                                    /note="assembly_fragment"
                                                                    120902. 156534
                                                                        /note="assembly_fragment"

```

ORIGIN

Query Match 80.0%; Score 398.4; DB 2; Length 156534;

Best Local Similarity 95.1%; Pred. No. 5.5e-88;
Matches 411; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```

QY 51 ACTCAGTACGAGCGCCGCTTCTGGTACGCTGGCAACCGGCGTCTGCTCTTCTCT 110
Db 96091 ACTCCACATGCCCTCTTCTTTTCATCCCAACCGCCACAGGCGTCTGCTCTTCTCT 96150

QY 111 CGAGAGGGCGGTGAGTCTTCAGTATGTTCCGCCAGCGCTCTTGCACCCCTTCTGGA 170
Db 96151 CGAGAGGGCGGTGAGTCTTCAGTATGTTCCGCCAGCGCTCTTGCACCCCTTCTGGA 96210

```

```

QY 171 CCAAGCGCCCAAGACTGCAGCCAGAGAGAGGGGCTCACTTATCTCGCGACCC 230
Db 96211 CCAAGCGCCCAAGACTGCAGCCAGAGAGAGGGGCTCACTTATCTCGCGACCC 96270

QY 231 ACTGCACAGCAGGCGCGCTCTCCAGACTTAATAATGATATACCACTAAGCTGAGGGG 290
Db 96271 ACTGCACAGCAGGCGCGCTCTCCAGACTTAATAATGATATACCACTAAGCTGAGGGG 96330

QY 291 ACCCAATCTGACTCTTCCCGGCTTGGACATCGCAGGCGGGAAGAGTGCCTGCCA 350
Db 96331 ACCCAATCTGACTCTTCCCGGCTTGGACATCGCAGGCGGGAAGAGTGCCTGCCA 96390

QY 351 GGCCTGGGCCAGAGAGACTCCAGGAAGGCACTGAGCGCTGCTGGCGGAGGCTCGAC 410
Db 96391 GGCCTGGGCCAGAGAGACTCCAGGAAGGCACTGAGCGCTGCTGGCGGAGGCTCGAC 96450

QY 411 ATCCGACAGCAGCAGGAAGTCTCTCGGCGATCTGTAATAAACCTTTTCTTTT 470
Db 96451 ATCCGACAGCAGCAGGAAGTCTCTCGGCGATCTGTAATAAACCTTTTCTTTT 96510

QY 471 GTTTTAAAAA 482
Db 96511 GTTTTAAAAA 96522

```

RESULT 7
AC012255
LOCUS
DEFINITION Homo sapiens chromosome 15 clone RP11-109D20 map 15, WORKING DRAFT
SEQUENCE, 22 unordered pieces.
AC012255
AC012255.4 GI:7684440
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens

REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karates,A., Klein,J., Lehoczy,J., Iley,C., Locke,K., Macdonald,P., Margus,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Tittrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
JOURNAL Direct Submission
REFERENCE 2 (bases 1 to 166937)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome 15, clone RP11-109D20
REFERENCE 1 (bases 1 to 166937)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome 15, clone RP11-109D20
REFERENCE 2 (bases 1 to 166937)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karates,A., Klein,J., Lehoczy,J., Iley,C., Locke,K., Macdonald,P., Margus,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Tittrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3737
Center clone name: 109 D 20
----- Summary Statistics
Sequencing vector: M13, M77815, 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 153800 bases at least Q40
Consensus quality: 160606 bases at least Q30
Consensus quality: 163275 bases at least Q20
Insert size: 164837; sum-of-contigs
Quality coverage: 4.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2087: contig of 2087 bp in length
* 2088 2187: gap of 100 bp
* 2188 3626: contig of 1439 bp in length
* 3627 3726: gap of 100 bp
* 3727 4845: contig of 1119 bp in length
* 4846 4945: gap of 100 bp
* 4946 6702: contig of 1757 bp in length
* 6703 6802: gap of 100 bp
* 6803 8326: contig of 1524 bp in length
* 8327 8426: gap of 100 bp
* 8427 10494: contig of 2068 bp in length
* 10495 10594: gap of 100 bp
* 10595 13215: contig of 2621 bp in length
* 13216 13315: gap of 100 bp
* 13316 16451: contig of 3136 bp in length
* 16452 16552: gap of 100 bp
* 16552 20843: contig of 4292 bp in length
* 20844 20943: gap of 100 bp
* 20944 24940: contig of 3997 bp in length
* 24941 25041: gap of 100 bp
* 25041 29901: contig of 4861 bp in length
* 29902 30001: gap of 100 bp
* 30002 34189: contig of 4188 bp in length
* 34190 34289: gap of 100 bp
* 34290 38386: contig of 4097 bp in length
* 38387 38486: gap of 100 bp
* 38487 45827: contig of 7341 bp in length
* 45828 45927: gap of 100 bp
* 45928 51729: contig of 5802 bp in length
* 51730 51829: gap of 100 bp
* 51830 60123: contig of 8294 bp in length
* 60124 60223: gap of 100 bp
* 60224 73532: contig of 13309 bp in length
* 73533 73632: gap of 100 bp
* 73633 87433: contig of 13801 bp in length
* 87434 87533: gap of 100 bp
* 87534 101708: contig of 14175 bp in length
* 101709 101808: gap of 100 bp
* 101809 121193: contig of 19385 bp in length
* 121194 121293: gap of 100 bp
* 121294 139051: contig of 17758 bp in length
* 139052 139151: gap of 100 bp
* 139152 166937: contig of 27786 bp in length.
location/Qualifiers

FEATURES
source
1.166937
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-109D20"
/clone_1lb="RPC1-11 Human Male BAC"
1.2087
misc_feature
/note="assembly_fragment"
2188.3626
misc_feature
/note="assembly_fragment"
3727.4845

/note="assembly_fragment"
4946.6702
/note="assembly_fragment"
6803.8326
/note="assembly_fragment"
8427.10494
/note="assembly_fragment"
10595.13215
/note="assembly_fragment"
13316.16451
/note="assembly_fragment"
16552.20843
/note="assembly_fragment"
20944.24940
/note="assembly_fragment"
25041.29901
/note="assembly_fragment"
30002.34189
/note="assembly_fragment"
34290.38386
/note="assembly_fragment"
38487.45827
/note="assembly_fragment"
45928.51729
/note="assembly_fragment"
51830.60123
/note="assembly_fragment"
60224.73532
/note="assembly_fragment"
73633.87433
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
87534.101708
/note="assembly_fragment"
101809.121193
/note="assembly_fragment"
121294.139051
/note="assembly_fragment"
139152.166937
/note="assembly_fragment"
clone_end:T7
vector_side:right"

ORIGIN

Query Match 80.0%; Score 398.4; DB 2; Length 166937;
Best Local Similarity 95.1%; Pred. No. 5.5e-88;
Matches 411; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 51 ACTCAGTACGAGCGCGCTTCTGGGTACGCTGGCAACCGCGTCTGTGCTCTTCT
Db 161043 ACTCACATGACCTCTCTTCTTTCGATGCCACCGCCACAGCGCTGTGCTCTTCT
QY 111 CCGAGGGGGCGGTGTGAGTCTCCAGTATGTTGGGCCCGCGCTCTTGCACCCCTTGTGA
Db 161103 CCGAGGGGGCGGTGTGAGTCTCCAGTATGTTGGGCCCGCGCTCTTGCACCCCTTGTGA
QY 171 CCAAAGCGCCAGGACTGACGACGAGAGAGAGGGGGCTCACCTTTATCTCTGGGACCC
Db 161163 CCAAAGCGCCAGGACTGACGACGAGAGAGAGGGGGCTCACCTTTATCTCTGGGACCC
QY 231 ACTGCACACAGCAGGCGGCTCTCCAGACTTAAATGTATCACCACCTAACCCTGTAGGGGG
Db 161223 ACTGCACACAGCAGGCGGCTCTCCAGACTTAAATGTATCACCACCTAACCCTGTAGGGGG
QY 291 ACCCAATCTGACTCTCTTCCCGCGCTTGGGACATCGCAGGCGCGGGAAGCAGTGCCTGCCCA
Db 161283 ACCCAATCTGACTCTCTTCCCGCGCTTGGGACATCGCAGGCGCGGGAAGCAGTGCCTGCCCA
QY 351 GGCCTGGGCGCAGAGAGCTCCAGGAAGGACACTGAGCGCTGTGGCGGAGGCGCTCGGAC
Db 161343 GGCCTGGGCGCAGAGAGCTCCAGGAAGGACACTGAGCGCTGTGGCGGAGGCGCTCGGAC

QY 411 ATCCGAGCACCAGGAAAGTCTCTGGGGCGCATCTGTAATAACCTTTTCTTTT 470
|||||
Db 161403 ATCCGAGCACCAGGAAAGTCTCTGGGGCGCATCTGTAATAACCTTTTCTTTT 161462
QY 471 GTTTTAAAAA 482
|||||
Db 161463 GTTTTAAAAA 161474

RESULT 8
AC087790
LOCUS 171444 bp DNA linear HTG 09-MAY-2001
DEFINITION Homo sapiens chromosome 15 clone RP11-276K9 map 15, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
AC087790
ACCESSION AC087790.2 GI:13357344
VERSION AC087790
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 171444)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-276K9
Unpublished
2 (bases 1 to 171444)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
Lehoczký,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
Marguis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Piere,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testafay,S., Theodore,J.,
Travers,M., Travis,N., Trigliio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT
Submitted (24-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2001 this sequence version replaced gi:12405501.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L12451
Center clone name: 276_K_9

Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 165906 bases at least Q40
Consensus quality: 168587 bases at least Q30
Consensus quality: 169557 bases at least Q20
Insert size: 18000; agarose-fp
Insert size: 170244; sum-of-contigs
Quality coverage: 5.7 in Q20 bases; agarose-fp
Quality coverage: 6.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
6091: contig of 6091 bp in length
6092 6191: gap of 100 bp
6192 7986: contig of 1795 bp in length
7987 8086: gap of 100 bp
8087 43939: contig of 35853 bp in length
43940 44039: gap of 100 bp
44040 45298: contig of 1259 bp in length
45299 45398: gap of 100 bp
45399 47595: contig of 2197 bp in length
47596 47695: gap of 100 bp
47696 52385: contig of 4690 bp in length
52386 52485: gap of 100 bp
52486 56310: contig of 3825 bp in length
56311 56410: gap of 100 bp
56411 71313: contig of 14903 bp in length
71314 90993: contig of 19580 bp in length
90994 91093: gap of 100 bp
91094 111459: contig of 20366 bp in length
111460 111559: gap of 100 bp
111560 137726: contig of 26167 bp in length
137727 137826: gap of 100 bp
137827 170053: contig of 32227 bp in length
170054 170153: gap of 100 bp
170154 171444: contig of 1291 bp in length.
location/Qualifiers
1. 171444
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-276K9"
/clone_lib="RPC1-11 Human Male BAC"
1. 6091
/note="assembly_fragment"
clone_end:SP6
vector_side:left
6192..7986
/note="assembly_fragment"
8087..43939
/note="assembly_fragment"
44040..45298
/note="assembly_fragment"
45399..47595
/note="assembly_fragment"
47696..52385
/note="assembly_fragment"
52486..56310
/note="assembly_fragment"
56411..71313
/note="assembly_fragment"
71414..90993
/note="assembly_fragment"
91094..111459
/note="assembly_fragment"
111560..137726
/note="assembly_fragment"
137827..170053
/note="assembly_fragment"
170154..171444
clone_end:T7
vector_side:right"

ORIGIN

Query Match 80.0%; Score 398.4; DB 2; Length 171444;
 Best Local Similarity 95.1%; Pred. No. 5.5e-88;
 Matches 411; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 51 ACTCAGTACGAGCGCCGCTCTGTGGTACGCTGGCAACCGCGCTCTGTGCTCTTCT 110
 143740 ACTCCACATGCTCTCTTCTTTCGATCCCAACCGCCACAGCGCTCTGTGCTCTTCT 143799
 QY 111 CGGAGGGGGCGCTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTCCGACCCCTTCTGGA 170
 143800 CGGAGGGGGCGCTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTCCGACCCCTTCTGGA 143859
 QY 171 CCAAGCGCCAGAGACTGCAGCCAGAGAGAGAGGGGGCTCACCTCTTATCCTCGCGACCC 230
 143860 CCAAGCGCCAGAGACTGCAGCCAGAGAGAGAGGGGGCTCACCTCTTATCCTCGCGACCC 143919
 QY 231 ACTGCACAAGCAGCGCGCTCTCCAGACTTAAATGTATCACCACTAACCTGTGAGGGGG 290
 143920 ACTGCACAAGCAGCGCGCTCTCCAGACTTAAATGTATCACCACTAACCTGTGAGGGGG 143979
 QY 291 ACCCAATCTGACTCTTCCCGCGCTTGGGACATCGCAGCGCGGAGAGCAGTGCCTCGGCA 350
 143980 ACCCAATCTGACTCTTCCCGCGCTTGGGACATCGCAGCGCGGAGAGCAGTGCCTCGGCA 144039
 QY 351 GGCTTGGGCGCAGAGAGCTCCAGAAAGGCACTGAGCGCTGCTGGCGAGCGCTCGGAC 410
 144040 GGCTTGGGCGCAGAGAGCTCCAGAAAGGCACTGAGCGCTGCTGGCGAGCGCTCGGAC 144099
 QY 411 ATCCGAGGCGACCGAGGAAAGTCTCTGGGCGATCTGTAATAAACCTTTTCTTTT 470
 144100 ATCCGAGGCGACCGAGGAAAGTCTCTGGGCGATCTGTAATAAACCTTTTCTTTT 144159
 QY 471 GTTTTTTAAAAA 482
 144160 GTTTTTTAAAAA 144171
 Db

RESULT 9
 AC091117 181312 bp DNA linear PRI 20-FEB-2002
 LOCUS Homo sapiens chromosome 15 clone RP11-109D20 map 15q15, complete
 DEFINITION

ACCESSION AC091117
 VERSION AC091117.5 GI:18767395
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 181312)
 Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
 Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
 Pate, D. and Hood, L.

TITLE
 JOURNAL Sequencing of human chromosome 15 D15S146-D15S117 region
 REFERENCE
 AUTHORS Unpublished
 2 (bases 1 to 181312)
 Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
 Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
 Pate, D. and Hood, L.

TITLE
 JOURNAL Direct Submission
 Submitted (30-MAR-2001) Multimegabase Sequencing Center, Institute
 for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
 98105, USA

REFERENCE
 AUTHORS 3 (bases 1 to 181312)
 Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
 Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
 Pate, D. and Hood, L.
 TITLE
 JOURNAL Direct Submission
 Submitted (13-FEB-2002) Multimegabase Sequencing Center, Institute
 for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
 98105, USA

REFERENCE
 AUTHORS 4 (bases 1 to 181312)
 Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
 Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
 Pate, D. and Hood, L.

TITLE
 JOURNAL Direct Submission
 Submitted (14-FEB-2002) Multimegabase Sequencing Center, Institute
 for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
 98105, USA

REFERENCE
 AUTHORS 5 (bases 1 to 181312)
 Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
 Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
 Pate, D. and Hood, L.

TITLE
 JOURNAL Direct Submission
 Submitted (20-FEB-2002) Multimegabase Sequencing Center, Institute
 for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
 98105, USA

COMMENT On Feb 20, 2002 this sequence version replaced gi:18653547.

----- Genome Center

Center: Multimegabase Sequencing Center

Center code: UMSC

Web site: http://chroma.mbt.washington.edu/msg_www

Contact: leerowen@systemsbiology.org

----- Summary Statistics

Sequencing vector: pUC18; L08752

Chemistry: Dye-terminator Big Dye; 90% of reads

Chemistry: Dye-primer Big Dye; 10% of reads

Assembly program: Phrap; version 0.990399

Note: Data from overlapping clone AC090888 [Drafting center: WIBR]
 was added for finishing. RP11-109D20 spans bases 1-174236 of this
 sequence. Bases 174237-179373 derive from clone RP11-276K9,
 AC087790 [Drafting center: WIBR] to establish the overlap with the
 next BAC in our tiling path. Bases 179373-181312 derive from
 finished clone CTD-2651B20, AC051619 [Drafting center: UMSC] in
 order to give overlap required for determining long range
 contiguity.

FEATURES

source

1. 181312

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="15"

/map="15q15"

/clone="RP11-109D20"

/clone_lib="RPCI human BAC library 11"

/note="This clone overlaps CTD-2014N11 AC090888 and
 RP11-276K9 AC087790. Data from overlapping BACs were added
 and the consensus sequence determined from RP11-109D20 to
 the extent possible."

1. 89137

/note="overlap with CTD-2014N11, AC090888"

36107. 36120

/note="low quality data"

40470. 40820

/note="sequence data generated from subcloned PCR product"

53030. 53480

/note="sequence data generated from subcloned PCR product"

86982. 86983

/note="low quality data"

95610. 96030

/note="sequence data generated from subcloned PCR product"

96830. 97420

/note="sequence data generated from subcloned PCR product"

98260. 98420

/note="single clone coverage"

98850. 98852

/note="low quality data"

100981. 101003

/note="low quality data"

174237. 179372

/note="overlap with RP11-276K9, AC087790. This sequence
 data was added to establish the overlap between the BACs

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature 179373..181312
 in the tiling path"
 /note="overlap with CTD-2651B20, AC051619. This data is
 added to give overlap for long range contiguity."

ORIGIN

Query Match 80.0%; Score 398.4; DB 9; Length 181312;
 Best Local Similarity 95.1%; Pred. No. 5.5e-88;
 Matches 411; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 51 ACTCAGTACGACGCGCCCTTCTGGGTACGCTGGCAACCGCGTCTGTGCTCTTCT 110
 Db 168343 ACTCCACATGCCCTCCTTCTTTTCATGCCACCGCCACAGGCGTCTGTGCTCTTCT 168402
 QY 111 CGGAGGGGGCGGTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTCCACCTTCTTGA 170
 Db 168403 CGGAGGGGGCGGTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTCCACCTTCTTGA 168462
 QY 171 CCAAGCGCCAGAGACTGCAGCCAGAGAGAGAGGGGCTCACCTCTTATCTCCGCGACCC 230
 Db 168463 CCAAGCGCCAGAGACTGCAGCCAGAGAGAGAGGGGCTCACCTCTTATCTCCGCGACCC 168522
 QY 231 ACTGCACAAGAGGCGCGCTCTCCAGACTTAAATGTATCATCACTTAACCTGTAGGGGG 290
 Db 168523 ACTGCACAAGAGGCGCGCTCTCCAGACTTAAATGTATCATCACTTAACCTGTAGGGGG 168582
 QY 291 ACCCAATCTGGACTCTTCCCGCGCTTGGAGCATCGAGCGCGCGAGAGCAGTGCCTCCCA 350
 Db 168583 ACCCAATCTGGACTCTTCCCGCGCTTGGAGCATCGAGCGCGCGAGAGCAGTGCCTCCCA 168642
 QY 351 GGCTTGGGCGGAGAGAGCTCCAGAAAGGCACTGAGCGCTGTGGCGGAGCGCTCCGAC 410
 Db 168643 GGCTTGGGCGGAGAGAGCTCCAGAAAGGCACTGAGCGCTGTGGCGGAGCGCTCCGAC 168702
 QY 411 ATCCGACAGCACGAGGAAAGTCTCTGGGCGATGTGAATAAACCTTTTCTTTT 470
 Db 168703 ATCCGACAGCACGAGGAAAGTCTCTGGGCGATGTGAATAAACCTTTTCTTTT 168762
 QY 471 GTTTTAAAAA 482
 Db 168763 GTTTTAAAAA 168774

RESULT 10

HS183H12F/c 217 bp DNA linear PRI 18-OCT-1995
 LOCUS HS183H12F
 DEFINITION H.sapiens Cpg island DNA genomic MseI fragment, clone 183h12,
 forward read cpj183h12.ftla.

ACCESSION 257552
 VERSION 257552.1 GI:1028783
 KEYWORDS Cpg island; genomic MseI fragment.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1
 AUTHORS Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
 TITLE Purification of Cpg islands using a methylated DNA binding column
 JOURNAL Nat. Genet. 6 (3), 236-244 (1994)

COMMENT
 MEDLINE 94282070
 PUBMED 8012384
 REFERENCE 2 (bases 1 to 217)
 AUTHORS Dodsworth,S.J., Huckle,E., Wilkinson,P. and Micklem,G.
 TITLE Direct Submision
 JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk

COMMENT
 Vectors: pGEM-5Zf(-)
 Clones are available from the UK MRC Human Genome Mapping Project
 Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
 http://www.hgmp.mrc.ac.uk/ for details
 or contact: biohelp@hgmp.mrc.ac.uk.

FEATURES

source 1..217
 Location/Qualifiers

/organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="183h12"
 /sex="male"
 /tissue_type="blood"
 /clone_lib="CGI-1"
 /dev_stage="adult"

ORIGIN

Query Match 43.6%; Score 217; DB 9; Length 217;
 Best Local Similarity 100.0%; Pred. No. 4.4e-43;
 Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 AAATGTATCACCACTAACCTGTGAGGGGACCAATCTGACTCTTCCCGCCTTGGGA 321
 Db 217 AAATGTATCACCACTAACCTGTGAGGGGACCAATCTGACTCTTCCCGCCTTGGGA 158
 QY 322 CATGCGAGCGCGGGAAGCAGTGCCTCCAGGCTGGGCCAGAGAGAGCTCCAGGAAGGCA 381
 Db 157 CATGCGAGCGCGGGAAGCAGTGCCTCCAGGCTGGGCCAGAGAGAGCTCCAGGAAGGCA 98
 QY 382 CTGAGCGCTGTGCGCGCGAGGCGCTCGGACATCCGACAGGCAACGGAAGTCTCTGAGG 441
 Db 97 CTGAGCGCTGTGCGCGCGAGGCGCTCGGACATCCGACAGGCAACGGAAGTCTCTGAGG 38
 QY 442 CGATCTGTAATAAACCTTTTCTTTCTTTTCTTTT 478
 Db 37 CGATCTGTAATAAACCTTTTCTTTCTTTTCTTTT 1

RESULT 11

HS183A12R 218 bp DNA linear PRI 19-OCT-1995
 LOCUS HS183A12R
 DEFINITION H.sapiens Cpg island DNA genomic MseI fragment, clone 183a12,
 reverse read cpj183a12.ftla.

ACCESSION Z59954.1 GI:1031867
 VERSION Z59954
 KEYWORDS Cpg island; genomic MseI fragment.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1
 AUTHORS Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
 TITLE Purification of Cpg islands using a methylated DNA binding column
 JOURNAL Nat. Genet. 6 (3), 236-244 (1994)

COMMENT
 MEDLINE 94282070
 PUBMED 8012384
 REFERENCE 2 (bases 1 to 218)
 AUTHORS Dodsworth,S.J., Huckle,E., Wilkinson,P. and Micklem,G.
 TITLE Direct Submision
 JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk

COMMENT
 Vectors: pGEM-5Zf(-)
 Clones are available from the UK MRC Human Genome Mapping Project
 Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
 http://www.hgmp.mrc.ac.uk/ for details
 or contact: biohelp@hgmp.mrc.ac.uk.

FEATURES
 source 1..218
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="183a12"
 /sex="male"
 /tissue_type="blood"
 /clone_lib="CGI-1"
 /dev_stage="adult"

ORIGIN

Query Match 40.1%; Score 199.8; DB 9; Length 218;
 Best Local Similarity 97.7%; Pred. No. 8.1e-39;

Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL MEDLINE 22388257
PUBMED 12477932
2 (bases 1 to 1354)
Strausberg, R.
Direct Submission
Submitted (03-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAK Plate: 58 Row: g Column: 17
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13385245.

FEATURES

source

1. 1354
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:35715 IMAGE:4989815"
/tissue_type="Colon, normal. 5 month old male mouse."
/clone_lib="NCT CGAP_C024"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1. 1354
/gene="9030623N16R1k"
/db_xref="LOCUSID:66811"
/db_xref="MGI:1914061"
129. 1091
/codon_start=1
/product="RIKEN cDNA 9030623N16"
/protein_id="AAH31111.1"
/db_xref="GI:21411408"
/db_xref="LOCUSID:66811"
/translation="MTAWDGLPYFPQPRHAASFVPLILVILVLSLAASFILPGIRGSRWFLVRLVLSLIFGAEIVAFHSGDWFVGRVWNTSYKAFSPRVQVHGLHVGLAGVITLRTGTPROQLNETIDNERTWRLNEDYTKYHALEKGLPDPVLYLAEKFTPSPCGLYHOYHLAGHYAAATMVAFCFWILANALLSPAPLGLYGLALTTGATLFGVFAFASISVPLCHFRILGSAVLTTPYGAFWLTATGILSLILGAAVVIHTRPSALRSFIDLIVKDCSNQAKGNSPLTNNPQHEQLKSPDLNITLL"

gene

CDS

ORIGIN

Query Match 32.2%; Score 160.6; DB 10; Length 1354;
Best Local Similarity 65.8%; Pred. No. 4,4e-29;
Matches 331; Conservative 0; Mismatches 154; Indels 18; Gaps 6;

QY 2 TCTACGCTGCCGCTCTGCCCCGCTCCGCTAGGCTCCTCCGCGCTCACCACCTCAGTACGA 61
DB 810 TCCAGCGTGCCGCTCTG-CCACTTCGCGCTGGGCTCCGCGCTCCTCAGCTTACTAG- 867
QY 62 GCGCGCCTTCTGGGTTCACGCTGGCAACCGCGCTCTGTGCTCTTCTCGAGAGGGCGG 121
DB 868 GCGCTCCTTTGGCTCAGCTGGCCACCGGCATCTGAGCTCTCTCGAGAGGGCGG 927
QY 122 TGGTAGTCTCCAGTATGTTGGGCCAGCGCTCTTCCACCTTCTGACCAAGCGCA 181
DB 928 TGGTAGTCTCCAGTATGTTGGGCCAGCGCTCTTCCACCTTCTGAGTGTCA 987
QY 182 AGGACT--GCAGCCAGAGAGAGGGGCTCAGCTCTTATCTCGGACCCACTGCACA 238
DB 988 AAGACTGTAGCAACCAAGGCTAAAGAACTCAGCTCTCAGCTCAACCAACCCGACACG 1047
QY 239 AGCAGCGCGCTCTCCAGACTTAAATGTATCACCACTAAGCTGTAGAGGGAGCCAAATC 298
DB 1048 AACAGTTGAAGAGTCCAGACTTAA--TATTAACCACTCTCTGTGAATAATAGTACTC 1104
QY 299 TGACCTCTTCCCGCTTGGGACATCGCAGCGGGAGAGTGCCTCGGAGGCTG- 357
DB 1105 CGATTTCTACCCCTCTTGGGACCCCATAGACTGAAAGTGTGTAAGAGCGGTGCC 1164
QY 358 GCCAGAGAGCTCCAGAGAGGCACTGAGCGCTGTGCGGAGGCGCTCGACATCCGCA 417
DB 1165 AGAGCATAGCTGAGTCTGAGAACTGTTGCTCCCTGCGGACGACAGAGGGGACACCCA 1224
QY 418 GGCACC-----AGGAAAGTCTCTCGGGGAGTGTGTAATAAATTTTCTT 468
DB 1225 TGTGCTTGTGCTTGAAGAAATGATTTCTCAGAGAACTGTAATAAATTTTGTTC 1284
QY 469 TTGTTTAAATAAATAAATAA 491
DB 1285 TTTTCTTCTTCAAAAAA 1307

RESULT 14
AL844566 229583 bp DNA linear ROD 11-APR-2003
LOCUS AL844566/c
DEFINITION Mouse DNA sequence from clone RP23-173H17 on chromosome 2, complete sequence.

ACCESSION AL844566
VERSION AL844566.8 GI:29823197
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 229583)
AUTHORS Pelan, S.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Apr 11, 2003 this sequence version replaced gi:23476718.

COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession


```
QY      1 CTCTAGCGTCCGCTCTGCCCCGCTCCCGCTAGGCTCTCCGCGCTCACCACCTCAGTACG 60
Db      694 CTCGAGCGTGCTCTTTGCCAGCT-CCGCTCGGCTCTCCGAACTCACCACCTCAGTACG 752
QY      61 AGCGCCGCTTCTGGGTACGCTGGCAACCGCGCTGTGCTCTTCCCTCGAGAGGGCC 120
Db      753 -GTGCTGCTTTTGATCAGCTGGCCACCGGCATCTGTGCTCTCCCTCGAGAGAGCG 811
QY      121 GTGTGAGTCTCCAGTATGTTCGCCAGCGCTCTTCGACCCCTTCTGACCAAGCGCC 180
Db      812 GTGTGAGTCTCCAGTATGTTCGCCAGCGCTCTTCGACCCCTTCTGAGGGAAGTATC 871
QY      181 AAGGACTGCAGCCAGAGAGAGGGGCTCACCCTT 216
Db      872 AAGGACTGCAGTATGTCAAGCAAAAGGCTCTCCATTT 907
```

Search completed: February 23, 2004, 21:57:54
Job time : 1969.03 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 19:33:52 ; Search time 201.357 Seconds

(without alignments)
10506.704 Million cell updates/sec

Title: US-09-936-456-3

Perfect score: 498
Sequence: 1 ctctagcgtgcgcgtctgcc.....aaaaaaaaataaagtcgacc 498

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

1: geneseqn1980s:*

2: geneseqn1990s:*

3: geneseqn2000s:*

4: geneseqn2001as:*

5: geneseqn2001bs:*

6: geneseqn2002s:*

7: geneseqn2003as:*

8: geneseqn2003bs:*

9: geneseqn2003cs:*

10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	498	100.0	498	3	AAA94624 Human CAS
2	465.4	93.5	1441	3	AAA94623 Human CAS
3	465.4	93.5	1474	6	ABL90535 Human pol
4	465.4	93.5	1491	9	ADD19228 Human CDN
5	453	91.0	1421	3	AAA96505 Human enco
6	421.8	84.7	1460	3	AAC95564 Human sec
7	421.8	84.7	1460	7	ABZ67235 Human sec
8	421.8	84.7	1460	7	ABZ73640 Secreted
9	421.8	84.7	1460	9	ADC20289 Human sec
10	398.4	80.0	2684	7	ADA53627 Human cod
11	398.4	80.0	5033	7	ABZ68115 Human sec
12	398.4	80.0	5033	7	ABZ74587 Secreted
13	398.4	80.0	5033	9	ADC21005 Human sec
14	46	9.2	3180	4	AAL17543 Human CDN
15	46	9.2	5049	4	AAL03901 Human rep
16	46	9.2	5049	4	AAL03900 Human rep
17	46	9.2	5049	5	AAB40308 Human encod
18	46	9.2	5049	5	AAS40309 DNA encod
19	42	8.4	536	9	ADB68842 Minority
20	39	7.8	536	9	ADB68842 Minority
21	39	7.8	3786	5	AAS72683 DNA encod
22	38.8	7.8	8503	4	AAK81406 Human imm
23	38.6	7.8	2000	7	ADA71938 Rice gene

24	36.4	7.3	1751	6	ABL90392 Human pol
25	36.4	7.3	2074	4	AAD12572 Human pro
26	36.4	7.3	2134	3	AAA93126 Human sec
27	36.4	7.3	2156	6	ABN59662 Novel hum
28	36.2	7.3	2776	4	AAK68475 Human imm
29	36.2	7.3	2776	4	AAK68476 Human imm
30	36.2	7.3	2776	4	AAL04043 Human rep
31	36.2	7.3	2776	4	AAL04045 Human rep
32	35.6	7.1	1810	6	AAZ56767 Human tra
33	35.6	7.1	6232	6	AAL42135 Human kru
34	35.6	7.1	6232	6	AAL42134 Human kru
35	35.6	7.1	47988	9	ADB86070 Streptomy
36	35.4	7.1	1464	6	ADB86070 Streptomy
37	35.4	7.1	2080	2	AAZ10646 Gene enco
38	35.4	7.1	2080	7	ADB99281 Human cod
39	35.4	7.1	42000	3	AAA63349 Streptomy
40	35.4	7.1	63164	3	AAA63348 Streptomy
41	35.2	7.1	949	6	ABN74081 Bovine em
42	35.2	7.1	1395	7	AAD55820 Micromono
43	35.2	7.1	60196	7	AAD55810 Micromono
44	35	7.0	273	6	ABN18150 Human ORF
45	35	7.0	1395	4	AAS54132 Pseudomon

ALIGNMENTS

RESULT 1
AAA94624
ID AAA94624 standard; DNA; 498 BP.
XX
AC AAA94624;
XX
DT 11-JAN-2001 (first entry)
XX
DE Human CASB618 EST.
XX
KW Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian;
KW colon; autoimmune disease; HLA_A0201; expressed sequence tag; EST; ss.
XX
OS Homo sapiens.
XX
PN WO200053748-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-EP002048.
XX
PR 11-MAR-1999; 99GB-00005607.
XX
PR 01-SEP-1999; 99GB-00020590.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Bruck CEM, Cassart J, Coche T, Vinals Y De BassolsC;
XX
DR WPI; 2000-572268/53.
XX
PT New human CASB618 polypeptide, useful as a vaccine for prophylactic and
PT therapeutic treatment of cancers, particularly ovarian or colon cancer,
PT autoimmune diseases and related conditions.
XX
PS Claim 32; Page 62; 76pp; English.
XX
CC The present sequence is an expressed sequence tag (EST) for human CASB618
CC protein. The gene for human CASB618 is thought to be located on
CC chromosome 15. CASB618 protein and epitopes of CASB618 protein (see
CC AAB26327 to AAB26399) are useful in diagnosing the occurrence of tumour
CC cells and in vaccines for prophylactic and therapeutic treatment of
CC cancers, particularly ovarian or colon cancer, autoimmune diseases and
CC related conditions
XX
SQ Sequence 498 BP; 97 A; 166 C; 137 G; 98 T; 0 U; 0 Other;

Query Match 100.0%; Score 498; DB 3; Length 498;
Best Local Similarity 100.0%; Pred. No. 4.5e-129;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCTAGCGTCCGCTCTGCGCCGCTCCGCTAGGCTCTCCGCGCTCACCCTAGTACG 60
DB 1 CTCTAGCGTCCGCTCTGCGCCGCTCCGCTAGGCTCTCCGCGCTCACCCTAGTACG 60
QY 61 AGCGCCGCTTCTGGGTACAGCTGGCAACCGCGCTCTGTGCTCTTCTCTGGAGGGGCC 120
DB 61 AGCGCCGCTTCTGGGTACAGCTGGCAACCGCGCTCTGTGCTCTTCTCTGGAGGGGCC 120
QY 121 GTGTGAGTCTCCAGTATGTTCCGGCCAGCGCTCTTCCGACCCCTTCTGACCAAGCGCC 180
DB 121 GTGTGAGTCTCCAGTATGTTCCGGCCAGCGCTCTTCCGACCCCTTCTGACCAAGCGCC 180
QY 181 AAGACTGACGCCAGAGAGAGAGGGGCTCACCCTTATCTCGCGGACCCACTGCACAAG 240
DB 181 AAGACTGACGCCAGAGAGAGAGGGGCTCACCCTTATCTCGCGGACCCACTGCACAAG 240
QY 241 CAGCGCGCTTCCAGACTTAAATGATACCACTAACCTGTGAGGGGAGCCCAATCTG 300
DB 241 CAGCGCGCTTCCAGACTTAAATGATACCACTAACCTGTGAGGGGAGCCCAATCTG 300
QY 301 GACTCCTTCCCGCTTGGGACATCGCAGCGCGGGAAGCAGTGCCTCGCAGGCTGGGCC 360
DB 301 GACTCCTTCCCGCTTGGGACATCGCAGCGCGGGAAGCAGTGCCTCGCAGGCTGGGCC 360
QY 361 AGGAGAGCTCCAGGAAGGCACTGAGCGCTGTGGCGCGAGCGCTCGACATCCGACAGC 420
DB 361 AGGAGAGCTCCAGGAAGGCACTGAGCGCTGTGGCGCGAGCGCTCGACATCCGACAGC 420
QY 421 ACCAGGGAAGTCTCTCTGGGGGAGTGTAAATAAACCTTTTCTTTGTTTTTAA 480
DB 421 ACCAGGGAAGTCTCTCTGGGGGAGTGTAAATAAACCTTTTCTTTGTTTTTAA 480
QY 481 AAAAAATATAAGTCGACC 498
DB 481 AAAAAATATAAGTCGACC 498

RESULT 2
AAA94623 standard; DNA; 1441 BP.
XX AAA94623;
AC AAA94623;
XX 11-JAN-2001 (first entry)
DT 11-JAN-2001 (first entry)
XX Human CASB618 coding sequence.
DE Human CASB618 coding sequence.
XX Epitope; human; CASB618; chromosome 15; tumour; vaccine; cancer; ovarian;
KW colon; autoimmune disease; HLA_A0201; ss.
XX Homo sapiens.
OS Homo sapiens.
XX Key Location/Qualifiers
FH CDS 259..1221
FT /tag= a
FT /product= "Human CASB618"
XX
XX WO200053748-A2.
XX
XX 14-SEP-2000.
PD 14-SEP-2000.
XX
XX 09-MAR-2000; 2000MO-EP002048.
PF 09-MAR-2000; 2000MO-EP002048.
XX
XX 11-MAR-1999; 99GB-00005607.
PR 01-SEP-1999; 99GB-00020590.
XX
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
PA
XX Bruck CEM, Cassart J, Coche T, Vinals Y De BassolsC,
PI

XX WPI; 2000-572268/53.
DR P-PSDB; AAB26325.
DR
XX New human CASB618 polypeptide, useful as a vaccine for prophylactic and
PT therapeutic treatment of cancers, particularly ovarian or colon cancer,
PT autoimmune diseases and related conditions.
XX
XX Claim 13; Page 61; 76pp; English.
XX
XX The present sequence is the coding sequence of human CASB618 protein. The
CC gene for human CASB618 is thought to be located on chromosome 15. The
CC protein encoded by the present sequence and epitopes of the CASB618
CC protein (see AAB26327 to AAB26399) are useful in diagnosing the occurrence
CC of tumour cells and in vaccines for prophylactic and therapeutic
CC treatment of cancers, particularly ovarian or colon cancer, autoimmune
CC diseases and related conditions
XX
SQ Sequence 1441 BP; 289 A; 466 C; 392 G; 294 T; 0 U; 0 Other;

Query Match 93.5%; Score 465.4; DB 3; Length 1441;
Best Local Similarity 99.4%; Pred. No. 8.9e-120;
Matches 488; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 1 CTCTAGCGTCCGCTCTGCGCCGCTCCGCTAGGCTCTCCGCGCTCACCCTAGTACG 60
DB 939 CTCTAGCGTCCGCTCTGCGCCGCTCCGCTAGGCTCTCCGCGCTCACCCTAGTACG 997
QY 61 AGCGCCGCTTCTGGGTACAGCTGGCAACCGCGCTCTGTGCTCTTCTCTGGAGGGGCC 120
DB 998 -GCGCCGCTTCTGGGTACAGCTGGCAACCGCGCTCTGTGCTCTTCTCTGGAGGGGCC 1056
QY 121 GTGTGAGTCTCCAGTATGTTCCGGCCAGCGCTCTTCCGACCCCTTCTGACCAAGCGCC 180
DB 1057 GTGTGAGTCTCCAGTATGTTCCGGCCAGCGCTCTTCCGACCCCTTCTGACCAAGCGCC 1116
QY 181 AAGACTGACGCCAGAGAGAGAGGGGCTCACCCTTATCTCGCGGACCCACTGCACAAG 240
DB 1117 AAGACTGACGCCAGAGAGAGAGGGGCTCACCCTTATCTCGCGGACCCACTGCACAAG 1176
QY 241 CAGCGCGCTTCCAGACTTAAATGATACCACTAACCTGTGAGGGGAGCCCAATCTG 300
DB 1177 CAGCGCGCTTCCAGACTTAAATGATACCACTAACCTGTGAGGGGAGCCCAATCTG 1236
QY 301 GACTCCTTCCCGCTTGGGACATCGCAGCGCGGGAAGCAGTGCCTCGCAGGCTGGGCC 360
DB 1237 GACTCCTTCCCGCTTGGGACATCGCAGCGCGGGAAGCAGTGCCTCGCAGGCTGGGCC 1296
QY 361 AGGAGAGCTCCAGGAAGGCACTGAGCGCTGTGGCGGAGCGCTCGACATCCGACAGC 420
DB 1297 AGGAGAGCTCCAGGAAGGCACTGAGCGCTGTGGCGGAGCGCTCGACATCCGACAGC 1356
QY 421 ACCAGGGAAGTCTCTCTGGGGGAGTGTAAATAAACCTTTTCTTTGTTTTTAA 480
DB 1357 ACCAGGGAAGTCTCTCTGGGGGAGTGTAAATAAACCTTTTCTTTGTTTTTAA 1416
QY 481 AAAAAATATAAA 491
DB 1417 AAAAAATATAAA 1427

RESULT 3
ABL90535
ID ABL90535 standard; cDNA; 1474 BP.
XX ABL90535;
AC ABL90535;
XX
XX 24-MAY-2002 (first entry)
DT 24-MAY-2002 (first entry)
XX
XX Human polynucleotide SEQ ID NO 1097.
DE Human polynucleotide SEQ ID NO 1097.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;

KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; gene; ss.
OS Homo sapiens.
XX WO200190304-A2.
XX 29-NOV-2001.
XX 18-MAY-2001; 2001WO-US016450.
XX 19-MAY-2000; 2000US-0205515P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Birse CE, Rosen CA;
XX MPI; 2002-122018/16.
XX P-PSDB; ABB90126.
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders.
XX
XX
XX Claim 4; SEQ ID NO 1097; 2081bp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 1474 BP; 292 A; 480 C; 404 G; 296 T; 0 U; 2 Other;
Query Match 93.5%; Score 465.4; DB 6; Length 1474;
Best Local Similarity 99.4%; Pred. No. 9e-120;
Matches 488; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 1 CTCTAGCGTCCGCTCTGCGCCGCTCCGCTAGGCTCTCCGCGCTCACCACCTCAGTACG 60
Db 965 CTCTAGCGTCCGCTCTGCGCCGCT-CCGCTAGGCTCTCCGCGCTCACCACCTCAGTACG 1023
QY 61 AGCGCCGCTTCTGGGTACCGTGGCAACCGCGCTCTGCTTCTCTCGAGGGGCC 120
Db 1024 -GCGCCGCTTCTGGGTACCGTGGCAACCGCGCTCTGCTTCTCTCGAGGGGCC 1082
QY 121 GTGTGAGTCTCCAGTATGTTCGGCCAGCGCTCTTCGACCCCTTCTGGACCAAGCGCC 180
Db 1083 GTGTGAGTCTCCAGTATGTTCGGCCAGCGCTCTTCGACCCCTTCTGGACCAAGCGCC 1142
QY 161 AAGGACTGAGCCAGAGAGAGGGGGTCACTTATCTCTGGGAGACCCACGACAAAG 240
Db 1143 AAGGACTGAGCCAGAGAGAGGGGGTCACTTATCTCTGGGAGACCCACGACAAAG 1202
QY 241 CAGGCGGCTCTCCAGAGCTTAAATGTATCACCACCTAGTGAAGGGGAGCCCAATCTG 300
Db 1203 CAGGCGGCTCTCCAGAGCTTAAATGTATCACCACCTAGTGAAGGGGAGCCCAATCTG 1262
QY 301 GACTCCTTCCCGCCTTGGGACATCGCAGGCGGGAAACAGTGCAGCCGAGGCTTGGGCC 360

Db 1263 GACTCCTTCCCGCCTTGGGACATCGCAGGCGGGAAAGAGTGCAGCCAGCGCTGGGCC 1322
QY 361 AGGAGAGCTCCAGAGAGGGCACTGAGCGCTGCTGCGCGAGGCGCTCGACATCCGAGGC 420
Db 1323 AGGAGAGCTCCAGAGAGGGCACTGAGCGCTGCTGCGCGAGGCGCTCGACATCCGAGGC 1382
QY 421 ACCAGGAAAGTCTCTGCGGCGATCTGTAATAAACCTTTTCTTTGTTTTTAA 480
Db 1383 ACCAGGAAAGTCTCTGCGGCGATCTGTAATAAACCTTTTCTTTGTTTTTAA 1442
QY 481 AAAAAATAAAA 491
Db 1443 AAAAAAAAAA 1453
RESULT 4
ADD19228
ID ADD19228 standard; cDNA; 1491 BP.
XX AC ADD19228;
XX
XX 15-JAN-2004 (first entry)
DE Human cDNA from secreted protein gene 45.
XX
XX human secreted protein; cytosstatic; antibacterial; virucide;
KW neuroprotective; gynaecological; gastrointestinal-Gen; cardiant;
KW cardiovascular-Gen; nephrologic; antiinflammatory; muscular-Gen;
KW respiratory-Gen; immunosuppressive; cerebroprotective; vasotropic;
KW noctropic; antiallergic; cancer; bacterial infection; viral infection;
KW neural disorder; immune system disorder; blood disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW inflammatory disorder; proliferative disorder; human; ss; gene.
XX
XX Homo sapiens.
XX
XX WO2003052377-A2.
XX
XX 26-JUN-2003.
XX
XX 06-NOV-2002; 2002WO-US035606.
XX
XX 07-NOV-2001; 2001US-0331046P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
PI MPI; 2003-533050/50.
XX P-PSDB; ADD19303.
XX
XX New isolated nucleic acids encoding signal transduction pathway component
PT polypeptides, useful for diagnosing, treating, and/or preventing
PT disorders, such as cancer, infections, cardiovascular and inflammatory
PT diseases.
XX
XX Claim 1; SEQ ID NO 55; 554bp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (cDNA)
CC encoding a human secreted protein, representing one of 85 novel genes.
CC Also included are recombinant vectors, host cells (expressing the
CC protein), the secreted proteins (including their fragments, epitopes and
CC homologues), an isolated antibody that binds specifically to the protein,
CC diagnosing a pathological condition or susceptibility to a pathological
CC condition (comprising determining the presence or absence of a mutation
CC in the nucleic acid and diagnosing a condition based on the presence or
CC absence of the mutation), diagnosing a pathological condition or
CC susceptibility to a pathological condition (comprising determining the
CC presence or amount of expression of the protein in a biological sample
CC and diagnosing a condition based on the presence or amount of expression
CC of the protein), preventing, treating or ameliorating a medical condition

CC by administering the nucleic acid or protein to a mammalian subject,
CC identifying a binding partner to the protein, the gene corresponding to
CC the cDNA sequence, and identifying an activity in a biological assay
CC (comprising expressing the nucleic acid in a cell, isolating the
CC supernatant, detecting an activity in a biological assay and identifying
CC the protein in the supernatant having the activity). The nucleic acids
CC and proteins display the following activities Cytostatic, antibacterial,
CC Virucide, Neuroprotective, Gynaecological, Gastrointestinal-Gen,
CC Cardiant, Cardiovascular-Gen, Nephrotropic, Antiinflammatory, Muscular-
CC Gen, Respiratory-Gen, Immunosuppressive, Cerebroprotective, Vasotropic,
CC Neurotropic, Antiallergic. The methods and compositions of the present
CC invention are useful for diagnosing, treating, preventing and/or
CC prognosticating disorders related to the novel polypeptides, such as
CC cancer, bacterial or viral infections, and neural, immune system, blood,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal, inflammatory or proliferative disorders (many examples of these
CC diseases and disorders are given in the specification). The present
CC sequence encodes a novel secreted protein of the invention.

SQ Sequence 1491 BP, 293 A; 485 C; 410 G; 302 T; 0 U; 1 Other;

Query Match 93.5%; Score 465.4; DB 9; Length 1491;
Best Local Similarity 99.4%; Pred. No. 9e-120;
Matches 488; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 1 CTCTAGCGTCCGCTCTGCCCCGCTCCGCTAGGCTCCTCCGCGCTCACCACTCAGTACG 60
Db 982 CTCTAGCGTCCGCTCTGCCCCGCT-CCGCTAGGCTCCTCCGCGCTCACCACTCAGTACG 1040
QY 61 AGCGCCGCTTCTGGGTACAGCTGGCAACCGCGCTCTGTGCTCTTCTCTGGAGGGGCC 120
Db 1041 -GCGCCGCTTCTGGGTACAGCTGGCAACCGCGCTCTGTGCTCTTCTCTGGAGGGGCC 1099
QY 121 GTGTGAGTCTCCAGTATGTTCGGCCCAAGCGCTCTTCCGACCTTCTGGACCAAGCGCC 180
Db 1100 GTGTGAGTCTCCAGTATGTTCGGCCCAAGCGCTCTTCCGACCTTCTGGACCAAGCGCC 1159
QY 181 AAGGACTGCAGCCAG 240
Db 1160 AAGGACTGCAGCCAG 1219
QY 241 CAGGCGGCTTCCAGACTTAAATGATATACCACTAACCTGTGAGGGGAGCCCAATCTG 300
Db 1220 CAGGCGGCTTCCAGACTTAAATGATATACCACTAACCTGTGAGGGGAGCCCAATCTG 1279
QY 301 GACTCCTTCCCGCTTGGGACATCGCAGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 1280 GACTCCTTCCCGCTTGGGACATCGCAGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1339
QY 361 AGGAGAGTCCAGGAAGGAGCACTGAGCGCTGTGGCGGAGGCTCGACATCCGACAGG 420
Db 1340 AGGAGAGTCCAGGAAGGAGCACTGAGCGCTGTGGCGGAGGCTCGACATCCGACAGG 1399
QY 421 ACCAGGAAAGTCTCTGGGGCGATCTGTAATAAACCTTTTCTTTTCTTTTAA 480
Db 1400 ACCAGGAAAGTCTCTGGGGCGATCTGTAATAAACCTTTTCTTTTCTTTTAA 1459
QY 481 AAAAAATATAA 491
Db 1460 AAAAAATAA 1470

RESULT 5
AAA96505 standard; cDNA; 1421 BP.

XX AAA96505;

DT 08-FEB-2001 (first entry)

DE cDNA encoding a human transmembrane protein.

XX Human; transmembrane protein; cell proliferation disorder; myeloma;

KW reproductive disorder; smooth muscle disorder; neurological disorder;
KW arteriosclerosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
KW allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;
KW Alzheimer's disease; Tourette's disorder; ss.

OS Homo sapiens.

XX Key Location/Qualifiers
FT CDS 267..1229
FT /*tag= a

XX WO200056891-A2.

XX 28-SEP-2000.

XX 22-MAR-2000; 2000WO-US007817.

XX 22-MAR-1999; 99US-0125537P.

XX 16-JUN-1999; 99US-0139565P.

XX (INCY-) INCYTE PHARM INC.

XX Yue H, Lal P, Tang YT, Hillman JL, Reddy R, Bandman O;
PI Baughn MR, Lu DM, Azimzai Y, Yang J;

XX WPI; 2000-579485/54.
DR P-PSDB; AAB18992.

PT New human transmembrane proteins are used to treat a disease or condition
PT associated with decreased expression of functional HTMP e.g. Tourette's
PT disorder, angina and leukemia.

XX Claim 4; Page 129; 130pp; English.

XX The present sequence encodes a human transmembrane proteins (HTMP).
CC Agonists and antagonists of the protein are used to treat a disease or
CC condition associated with overexpression of the protein. Diseases and
CC conditions which can be treated include cell proliferative,
CC immunological, reproductive, smooth muscle and neurological disorders
CC e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency
CC syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,
CC stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The
CC polynucleotides may be used to detect and quantify gene expression in
CC biopsied tissues where protein expression may be correlated with disease
CC e.g. to determine absence, presence or excess expression of HTMP or to
CC monitor regulation of HTMP expression during therapeutic intervention

SQ Sequence 1421 BP, 264 A; 468 C; 395 G; 294 T; 0 U; 0 Other;

Query Match 91.0%; Score 453; DB 3; Length 1421;
Best Local Similarity 99.6%; Pred. No. 2.6e-116;
Matches 475; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 CTCTAGCGTCCGCTCTGCCCCGCTCCGCTAGGCTCCTCCGCGCTCACCACTCAGTACG 60
Db 947 CTCTAGCGTCCGCTCTGCCCCGCT-CCGCTAGGCTCCTCCGCGCTCACCACTCAGTACG 1005
QY 61 AGCGCCGCTTCTGGGTACAGCTGGCAACCGCGCTGTGCTCTTCTCTGGAGGGGCC 120
Db 1006 -GCGCCGCTTCTGGGTACAGCTGGCAACCGCGCTGTGCTCTTCTCTGGAGGGGCC 1064
QY 121 GTGTGAGTCTCCAGTATGTTCGGCCCAAGCGCTCTTCCGACCTTCTGGACCAAGCGCC 180
Db 1065 GTGTGAGTCTCCAGTATGTTCGGCCCAAGCGCTCTTCCGACCTTCTGGACCAAGCGCC 1124
QY 181 AAGGACTGCAGCCAG 240
Db 1125 AAGGACTGCAGCCAG 1184
QY 241 CAGGCGGCTTCCAGACTTAAATGATATACCACTAACCTGTGAGGGGAGCCCAATCTG 300
Db 1185 CAGGCGGCTTCCAGACTTAAATGATATACCACTAACCTGTGAGGGGAGCCCAATCTG 1244

QY 301 GACTCCTTCCCGCCTTGGGACATGACAGGCCGGGAGACAGTGTCCCGCCAGGCTGAGCC 360
DB 1245 GACTCCTTCCCGCCTTGGGACATGACAGGCCGGGAGACAGTGTCCCGCCAGGCTGAGCC 1304
QY 361 AGGAGAGCTCCAGAAAGGCACTGAGCGCTGCTGCGCGGAGGCTTCGGACATCCGAGGC 420
DB 1305 AGGAGAGCTCCAGAAAGGCACTGAGCGCTGCTGCGCGGAGGCTTCGGACATCCGAGGC 1364
QY 421 ACCAGGGAAGTCTCCTGGGGCGATCTGTAATAAACCCTTTTCTTTGTTTTT 477
DB 1365 ACCAGGGAAGTCTCCTGGGGCGATCTGTAATAAACCCTTTTCTTTGTTTTT 1421
RESULT 6
AAC95564 ID AAC95564 standard, cDNA; 1460 BP.
AAC95564; AC XX
XX 21-FEB-2001 (first entry)
DE Human secreted protein gene 44 SEQ ID NO:54.
XX Human, secreted protein; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antitumor; antiparasitic; anticonvulsant; antibacterial;
KW antifungal; antiparasitic; cardiant; cancer; immune disease; allergy;
KW cardiovascular disorder; wound healing; infection; neurological disease;
KW ss.
XX Homo sapiens.
OS
PN WO200061596-A1.
XX 19-OCT-2000.
PD
XX 06-APR-2000; 2000WO-US008983.
PF
XX 09-APR-1999; 99US-0128703P.
PR 14-JAN-2000; 2000US-0176068P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-611865/58.
DR P-PSDB; AAB52055.
DR
XX
PT Fifty nucleic acid molecules encoding human secreted proteins, useful in
PT the prevention, treatment and diagnosis of cancer, immune disorders,
PT cardiovascular disorders and neurological diseases.
XX
XX Claim 1; Page 443; 505pp; English.
PS
XX Polynucleotide sequences AAC95521 - AAC95570 represent cDNA encoding 50
CC human secreted proteins AAB52012 - AAB52161. Sequences AAB52062 -
CC AAB52103 represent alternative polypeptides encoded by the genes, and
CC amino acid sequences with which they share homology. The genes and
CC proteins have activities dependent on the tissues and cells in which they
CC are expressed. Examples of their activities include cytostatic;
CC immunosuppressive; nootropic; neuroprotective; antiviral; antiallergic;
CC hepatotropic; antidiabetic; antiparasitic; anticonvulsant; antitumor;
CC anticonvulsant; antibacterial; antifungal; antiparasitic; and cardiant.
CC The secreted proteins, polypeptides, antagonists and agonists may be
CC useful in treating, preventing and/or diagnosing diseases and disorders
CC such as cancer, particularly breast and ovarian cancer, and other cancers
CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,
CC liver, lung, or urogenital. Immune disorders such as Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as

CC cerebral anoxia and epilepsy; and infectious diseases such as viral, the
CC bacterial, fungal and parasitic infections may also be treated using the
CC proteins and polynucleotides of the invention. Sequences AAC95512 -
CC AAC95520 and AAB52011 are used in the isolation and characterisation of
CC the proteins and polynucleotides of the invention
XX
SQ Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;
Query Match 84.7%; Score 421.8; DB 3; Length 1460;
Best Local Similarity .97.6%; Pred. No. 1.4e-107;
Matches 481; Conservative 0; Mismatches 7; Indels 5; Gaps 5;
QY 1 CTCTAGCGTGGCGCTCTGCGCGCTCCGCGCTAGGCTCTCCGCGCTCACCACTCAGTACG 60
DB 960 CTCTAGCGTGGCGCTCTGCGCGCT-CCGCGTAGGCTCTCCGCGCTCACCACTCAGTACG 1018
QY 61 AGCGCGCCTTCTGGGTGACGCTGACCAACCGCGCTGCTGCTCTTCTCTCGAGGGGCC 120
DB 1019 -GCGCGCCTTCTGGGTGACGCTGACCAACCGCGCTGCTGCTCTTCTCTCGAGGGGCC 1077
QY 121 GTGTGAGTCTCCAGTATGTTGCGGCCAGCGCTCTTGCACCCCTTCTGACCAAGCGCC 180
DB 1078 GTGTGAGTCTCCAGTATGTTGCGGCCAGCGCTCTTGCACCCCTTCTGACCAAGCGCC 1137
QY 181 AAGGACTGC-AGCCAGAGAGAGGGGCTCACTCTTATCTCTGGGCACTGACGAA 239
DB 1138 AAGGACTGCAGAGAGAGAGGGGCTCACTCTTATCTCT-GGGACCCACTGACGAA 1196
QY 240 GCAGGCGCGCTCTCCAGAC-TTAAATGTATCACCACCTAACCCTGTGAGGGGAGCCCAATC 298
DB 1197 GCAGGCGCGCTCTCCAGACCTTAAATGTATCACCACCTAACCCTGTGAGGGGAGCCCAATC 1256
QY 299 TGGACTCTTCCCGCGCTTGGGACATCGCAGGCCGGGAAAGCAGTGTCCCGCCAGGCTGCG 358
DB 1257 TGGACTCTTCCCGCGCTTGGGACATCGCAGGCCGGGAAAGCAGTGTCCCGCCAGGCTGCG 1316
QY 359 CCAGGAGAGCTCCAGAGAGGCACTGAGCGCTGCTGCGCGAGGCGCTCGACATCCGAG 418
DB 1317 CCAGGAGAGCTCCAGAGAGGCACTGAGCGCTGCTGCGCGAGGCGCTCGACATCCGAG 1376
QY 419 GCACCAAGGAAAGTCTCTGCGGCGATCTGTAATAAACCCTTTTCTTTGTTTTTA 478
DB 1377 GCACCAAGGAAAGTCTCTGCGGCGATCTGTAATAAACCCTTTTCTTTGTTTTTA 1436
QY 479 AAAAAAAAAATAAA 491
DB 1437 AAAAAAAAAAAAAA 1449
RESULT 7
ABZ67235 ID ABZ67235 standard, cDNA; 1460 BP.
XX
XX ABZ67235;
AC
XX
DT 26-MAR-2003 (first entry)
XX
DE Human secreted protein encoding cDNA SEQ ID NO 355.
XX
KW Human, secreted protein; nootropic; neuroprotective; cytostatic;
KW virucide; dermatological; immunosuppressive; antitubercular; anti-HIV;
KW vulnary; antibacterial; antiparkinsonian; antisticking; antianemic;
KW antitubercular; cancer; antirheumatic; hepatotropic; cerebroprotective;
KW antiinflammatory; antiallergic; antidiabetic; antitumor; anticonvulsant;
KW antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;
KW cardiovascular disorder; neurological disease; nephrotropic;
KW gene therapy; gene; chromosome 9p21; ds.
XX
OS Homo sapiens.
XX
PN WO20027186-A2.
XX
PD 03-OCT-2002.

XX 26-MAR-2002; 2002WO-US009188.
PF
XX
PR 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-040583/03.
DR P-PSDB; ABP99814.
XX
PT New human secreted proteins encoded by genes contained in cDNA clones
PT (e.g. HGCAC19), useful for preventing, treating or diagnosing e.g. AIDS,
PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
PT West Nile fever.
XX
PS Claim 7; Page 1353; 2423pp; English.
XX
CC The invention relates to novel human genes (ABZ66891-ABZ68209) and the
CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections
XX
SQ Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;

Query Match 84.7%; Score 421.8; DB 7; Length 1460;
Best Local Similarity 97.6%; Pred. No. 1.4e-107;
Matches 481; Conservative 0; Mismatches 7; Indels 5; Gaps 5;

QY 1 CTCTAGCGTGCCGCTCTGCGCCGCTCCCGCTAGGCTCTCCGCGCTCACCAGTACG 60
DB 960 CTCTAGCGTGCCGCTCTGCGCCGCTCCCGCTAGGCTCTCCGCGCTCACCAGTACG 1018

QY 61 AGCGCCGCTTGTGGTACGCTGGCAACCGGCGTCTGTGCTCTTCTCGAGGGGCC 120
DB 1019 -GCGCGCGCTTGTGGTACGCTGGCAACCGGCGTCTGTGCTCTTCTCGAGGGGCC 1077

QY 121 GTGTGAGTCTCCAGATATGTTGGGCCAGCGCTCTTTCGACCCCTTCTGACCAAGCGCC 180
DB 1078 GTGTGAGTCTCCAGATATGTTGGGCCAGCGCTCTTTCGACCCCTTCTGACCAAGCGCC 1137

QY 181 AAGGACTGC-AGCCAGAGAGAGGGGGCTCACCCTTATCTCTGGCGACCACTGCACAA 239
DB 1138 AAGGACTGCAGAGAGAGGGGGCTCACCCTTATCTCTGGCGACCACTGCACAA 1196

QY 240 GCAGCGCGCTCTCCAGAC-TTAAATGTATCACCAGTACCTGTGAGGGGACCAATC 298
DB 1197 GCAGCGCGCTCTCCAGACTTAAATGTATCACCAGTACCTGTGAGGGGACCAATC 1256

QY 299 TGGACTCTTCCCGCTTGGGACATCGCAGCGCCGGAAGCAGTGCCTCGAGGCTGGG 358
DB 1257 TGGACTCTTCCCGCTTGGGACATCGCAGCGCCGGAAGCAGTGCCTCGAGGCTGGG 1316

QY 359 CCAGGAGAGCTCCAGAGAGGCGACTGAGCGCTGTGGCGGAGCGCTCGAGATCCGAG 418
DB 1317 CCAGGAGAGCTCCAGAGAGGCGACTGAGCGCTGTGGCGGAGCGCTCGAGATCCGAG 1376

QY 419 GCACGAGGAAAGTCTCTGGGGGCGATCTGTAATAACCTTTTCTTTGTTT 478

DB 1377 GCACGAGGAAAGTCTCTGGGGGCGATCTGTAATAACCTTTTCTTTGTTT 1436
QY 479 AAAAAAAAAATAAA 491
DB 1437 AAAAAAAAAAAAAA 1449

RESULT 8
ABZ73640
ID ABZ73640 standard; cDNA; 1460 BP.
XX
AC ABZ73640;
XX
DT 12-MAY-2003 (first entry)
XX
DE Secreted protein-encoding gene 360 cDNA clone HUFCL31, SEQ ID NO:370.
XX
KW Human; secreted protein; cancer; tumour; hyperproliferative disorder;
KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;
KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
KW drug screening; chromosome identification; chromosome mapping;
KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
KW antianaemic; vulnery; chromosome 9p21; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200277013-A2.
XX
PD 03-OCT-2002.
XX
PE 26-MAR-2002; 2002WO-US009370.
XX
PR 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-040578/03.
DR P-PSDB; ABR01306.
XX
PT New human secreted proteins and nucleic acids, useful for detecting or
PT treating cancer or other hyperproliferative disorders, autoimmune
PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
XX
PS Claim 21; Page 1345; 2474pp; English.
XX
CC ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
CC invention also encompasses antibodies specific for the secreted proteins,
CC the use of the secreted proteins in drug screening and recombinant
CC vectors and host cells comprising a nucleic acid of the invention. The
CC secreted proteins are thought to be involved in biological activities
CC associated with cellular signalling, cellular differentiation, cell
CC migration, prohormone activation and neurotransmitter activity. The
CC secreted proteins, nucleic acids encoding them, antibodies or antibody
CC fragments specific for the secreted proteins, and modulators of protein
CC activity are useful for diagnosing or treating cancers or other
CC hyperproliferative disorders. Additionally, the secreted proteins and
CC their nucleic acids may also be used in the treatment of autoimmune
CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
CC wound healing. Nucleic acids of the invention may be used for chromosome
CC identification, chromosome mapping, in gene therapy, for identifying
CC individuals from minute biological samples, as hybridisation probes, and
CC as molecular weight markers. The present sequence represents a human
CC secreted protein-encoding cDNA clone of the invention
XX
SQ Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;

Query Match 84.7%; Score 421.8; DB 7; Length 1460;
Best Local Similarity 97.6%; Pred. No. 1.4e-107;
Matches 481; Conservative 0; Mismatches 7; Indels 5; Gaps 5;

QY 1 CTCTAGCGTGCCTCTGCGGCTCCCGCTAGGCTCTCCGCGCTCACCCTCAGTACG 60
DB 960 CTCTAGCGTGCCTCTGCGGCTCCCGCTAGGCTCTCCGCGCTCACCCTCAGTACG 1018
QY 61 AGCGCCGCTTCTGGGTACCGCTGCGCAACCGCGCTCTGTGCTCTTCTCGAGGGGCC 120
DB 1019 -GCGCCGCTTCTGGGTACCGCTGCGCAACCGCGCTCTGTGCTCTTCTCGAGGGGCC 1077
QY 121 GTGTGAGTCTCCAGTATGTTCGGCCAGCGCTCTTCCGACCCCTTCTGACCAAGCGCC 180
DB 1078 GTGTGAGTCTCCAGTATGTTCGGCCAGCGCTCTTCCGACCCCTTCTGACCAAGCGCC 1137
QY 181 AAGGACTGC-AGCCAGAGAGAGAGGGGGCTCACCTTTATCTCGCGCAGCCCACTGCACAA 239
DB 1138 AAGGACTGCAGAGCCAGAGAGGGGGCTCACCTTTATCT-GGGGAGCCCACTGCACAA 1196
QY 240 GCAGGCGCGCTCTCCAGAC-TTAAATGTATCACCACTAACCTGTGAGGGGAGCCCAATC 298
DB 1197 GCAGGCGCGCTTCCAGACTTTAAATGTATCACCACTAACCTGTGAGGGGAGCCCAATC 1256
QY 299 TGGACTCTTCCCGCTTGGGACATCGCAGGCGGGAAGCAGTGCCCGCCAGGCTGGG 358
DB 1257 TGGACTCTTCCCGCTTGGGACATCGCAGGCGGGAAGCAGTGCCCGCCAGGCTGGG 1316
QY 359 CCAGAGAGCTCCAGAGAGGAGGAGCTGAGCGCTGCGCGCGCTCGACATCCGACG 418
DB 1317 CCAGAGAGCTCCAGAGAGGAGGAGCTGAGCGCTGCGCGCGCTCGACATCCGACG 1376
QY 419 GCACCAAGGAAAGTCTCTGGGGGAGTGTAAATAAACCCTTTTCTTTGTTTTTA 478
DB 1377 GCACCAAGGAAAGTCTCTGGGGGAGTGTAAATAAACCCTTTTCTTTGTTTTTA 1436
QY 479 AAAAAAAAAATAAAA 491
DB 1437 AAAAAAAAAAAAAA 1449

RESULT 9
ADCC20289 standard; DNA; 1460 BP.

AC ADCC20289;
DT 18-DEC-2003 (first entry)
DE Human secreted protein coding sequence #228.
XX

KM gene therapy; human; secreted protein; haemopoietic disorder;
KM haematological disorder; anaemia; haemophilia; inflammatory disorder;
KM inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
KM leukaemia; wound healing; epithelial cell proliferation disorder;
KM immune disorder; autoimmune disorder; asthmatic disorder;
KM cardiovascular disorder; atherosclerosis; myocarditis;
KM infectious disease; HIV; AIDS; endocrine disorder; diabetes;
KM gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200292787-A2.
XX
PD 21-NOV-2002.
XX
PF 26-MAR-2002; 2002WO-US009257.
XX
PR 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.

XX
PI Rosen CA, Ruben SM;
XX WPI; 2003-129287/12.
DR
XX
PT New human secreted proteins and nucleic acid molecules, useful for
PT preparing a diagnostic or pharmaceutical composition for diagnosing,
PT preventing or treating hematopoietic or hematologic disorders, e.g.
PT anemia or hemophilia.
XX
PS Claim 1; SEQ ID NO 238; 1512bp; English.
XX
CC The invention comprises the amino acid and coding sequences of human
CC secreted proteins. The DNA and protein sequences of the invention are
CC useful for detecting, preventing, diagnosing, prognosticating, treating
CC or ameliorating: haematopoietic or hematological disorders (e.g. anaemia
CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease
CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
CC wound healing and disorders of epithelial cell proliferation; immune
CC disorders (e.g. autoimmune disorders and asthmatic disorders);
CC cardiovascular disease (e.g. atherosclerosis and myocarditis);
CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
CC and gastrointestinal disorders (e.g. duodenal ulcers and
CC gastroenteritis). The present DNA sequence encodes a human secreted
CC protein of the invention.
XX
SQ Sequence 1460 BP; 292 A; 472 C; 398 G; 298 T; 0 U; 0 Other;

Query Match 84.7%; Score 421.8; DB 9; Length 1460;
Best Local Similarity 97.6%; Pred. No. 1.4e-107;
Matches 481; Conservative 0; Mismatches 7; Indels 5; Gaps 5;

QY 1 CTCTAGCGTGCCTCTGCGGCTCCCGCTAGGCTCTCCGCGCTCACCCTCAGTACG 60
DB 960 CTCTAGCGTGCCTCTGCGGCTCCCGCTAGGCTCTCCGCGCTCACCCTCAGTACG 1018
QY 61 AGCGCCGCTTCTGGGTACCGCTGCGCAACCGCGCTCTGTGCTCTTCTCGAGGGGCC 120
DB 1019 -GCGCCGCTTCTGGGTACCGCTGCGCAACCGCGCTCTGTGCTCTTCTCGAGGGGCC 1077
QY 121 GTGTGAGTCTCCAGTATGTTCGGCCAGCGCTCTTCCGACCCCTTCTGACCAAGCGCC 180
DB 1078 GTGTGAGTCTCCAGTATGTTCGGCCAGCGCTCTTCCGACCCCTTCTGACCAAGCGCC 1137
QY 181 AAGGACTGC-AGCCAGAGAGAGGGGGCTCACCTTTATCTCGCGCAGCCCACTGCACAA 239
DB 1138 AAGGACTGCAGAGCCAGAGAGGGGGCTCACCTTTATCT-GGGGAGCCCACTGCACAA 1196
QY 240 GCAGGCGCGCTCTCCAGAC-TTAAATGTATCACCACTAACCTGTGAGGGGAGCCCAATC 298
DB 1197 GCAGGCGCGCTTCCAGACTTTAAATGTATCACCACTAACCTGTGAGGGGAGCCCAATC 1256
QY 299 TGGACTCTTCCCGCTTGGGACATCGCAGGCGGGAAGCAGTGCCCGCCAGGCTGGG 358
DB 1257 TGGACTCTTCCCGCTTGGGACATCGCAGGCGGGAAGCAGTGCCCGCCAGGCTGGG 1316
QY 359 CCAGAGAGCTCCAGAGAGGAGGAGCTGAGCGCTGCGCGAGGCGCTCGACATCCGACG 418
DB 1317 CCAGAGAGCTCCAGAGAGGAGGAGCTGAGCGCTGCGCGAGGCGCTCGACATCCGACG 1376
QY 419 GCACCAAGGAAAGTCTCTGGGGGAGTGTAAATAAACCCTTTTCTTTGTTTTTA 478
DB 1377 GCACCAAGGAAAGTCTCTGGGGGAGTGTAAATAAACCCTTTTCTTTGTTTTTA 1436
QY 479 AAAAAAAAAATAAAA 491
DB 1437 AAAAAAAAAAAAAA 1449

RESULT 10
ADA53627/c
ID ADA53627 standard; cDNA; 2684 BP.
XX

AC ADA53627;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human coding sequence, SEQ ID 1195.
XX
KM Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KM Gene Therapy; human; secretory protein; membrane proteins; cancer;
KM inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS Homo sapiens.
XX
PN EP1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
XX
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
XX
DR WPI; 2003-395539/38.
DR
DR P-PSDB; ADA55266.
XX
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 1; SEQ ID NO 1195; 205pp; English.
XX
XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 2684 BP; 599 A; 785 C; 684 G; 616 T; 0 U; 0 Other;

Query Match 80.0%; Score 398.4; DB 7; Length 2684;
Best Local Similarity 95.1%; Pred. No. 6.1e-101;
Matches 411; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 51 ACTCAGTACGAGCGCGCCTTCTGGGTACGCTGGCAACCGGCTCTGTGCTTCTTCT 110
DB 2383 ACTCCACATGCCCTCTTCTTCTTCGATCCCGCCACAGCGGCTCTGTGCTTCTTCT 2324

QY 111 CGAGGGGCGGTGTGAGTCTCCAGTATGTTGGGCCAGCGCTCTTGCACCCCTTCTGA 170
DB 2323 CGAGGGGCGGTGTGAGTCTCCAGTATGTTGGGCCAGCGCTCTTGCACCCCTTCTGA 2264

QY 171 CCAAGCGCCAAAGACTGACAGCCAGAGAGAGGGGCTCACCCTTTATCCTTGGCGACCC 230
DB 2263 CCAAGCGCCAAAGACTGACAGCCAGAGAGAGGGGCTCACCCTTTATCCTTGGCGACCC 2204

QY 231 ACTGCACAAGCAGGGCGGCTCTCCCAAGCTTAAATGTATCAACCACTAACCTGTGAGGGGG 290
DB 2203 ACTGCACAAGCAGGGCGGCTCTCCCAAGCTTAAATGTATCAACCACTAACCTGTGAGGGGG 2144

QY 291 ACCCAATCTGGAATCTCTTCCCGCGCTTGGGACATGCAAGGCCGGGAAGCAATGCCGCCCA 350
DB 2143 ACCCAATCTGGAATCTCTTCCCGCGCTTGGGACATGCAAGGCCGGGAAGCAATGCCGCCCA 2084

QY 351 GGCCTGGGCCAGAGAGCTCCAGGAAGGACACTGAGCGCTGCTGGCGCGAGGCTCGGAC 410
DB 2083 GGCCTGGGCCAGAGAGCTCCAGGAAGGACACTGAGCGCTGCTGGCGCGAGGCTCGGAC 2024

QY 411 ATCCGACGACACCGAGGAAAGTCTCTGGGGCGATCTGTAATTAACCTTTTCTTTT 470
DB 2023 ATCCGACGACACCGAGGAAAGTCTCTGGGGCGATCTGTAATTAACCTTTTCTTTT 1964

QY 471 GTTTTAAAAA 482
DB 1963 GTTTTAAAAA 1952

RESULT 11
ABZ68115
ID ABZ68115 standard; DNA; 5033 BP.
XX
AC ABZ68115;
XX
DT 26-MAR-2003 (first entry)
XX
DE Human secreted protein encoding genomic DNA SEQ ID NO 1638.
XX
XX
KM Human; secreted protein; nootropic; neuroprotective; cyostatic;
KM virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KM vulnery; antibacterial; antiparkinsonian; antischling; antianaemic;
KM antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
KM antiinflammatory; antiallergic; antidiabetic; antidiabetic; anticonvulsant;
KM antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine;
KM cardiovascular disorder; neurological disease; nephrotropic;
KM gene therapy; gene; ds.
XX
XX
OS Homo sapiens.
XX
XX
PN WO200277186-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-US009188.
XX
PR 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-040583/03.
XX
XX
PT New human secreted proteins encoded by genes contained in cDNA clones
PT (e.g. HGCA19), useful for preventing, treating or diagnosing e.g. AIDS,
PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
PT West Nile fever.
XX
PS Disclosure; Page 2263-2264; 2423pp; English.
XX
XX
CC The invention relates to novel human genes (ABZ66891-ABZ68209) and the
CC encoded secreted proteins (ABP9470-ABP9872) useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections
XX
SQ Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;

Query Match 80.0%; Score 398.4; DB 7; Length 5033;
Best Local Similarity 95.1%; Pred. No. 7.9e-101;

Matches 411; Conservative 0; Mismatches 21; Indels 0; Gaps 0;									
QY	51	ACTCAGTAGCAGCGCCGCTTCTGGGTCAACGCTGCAACCGCGCTCTGTCCTTCT	110						
Db	3340	ACTCCACATGCCCCCTCTTCTTTCATCCCGCCACAGGCGTCTGTCTCTTCT	3399						
QY	111	CGAGGGGCGGTGTAGTCTCCAGTATGTTCCGCCACGCGCTCTTCCGACCTTCTGA	170						
Db	3400	CGAGGGGCGGTGTAGTCTCCAGTATGTTCCGCCACGCGCTCTTCCGACCTTCTGA	3459						
QY	171	CCAAAGCGCCAGGACTGCAGCCAGAGAGAGAGGGGCTCACTCTTATCTCGGACCC	230						
Db	3460	CCAAAGCGCCAGGACTGCAGCCAGAGAGAGAGGGGCTCACTCTTATCTCGGACCC	3519						
QY	231	ACTGCAAGCAGCGCGCTCTCCAGACTTAAATGTATCACCACTAACCTGTAGGGG	290						
Db	3520	ACTGCAAGCAGCGCGCTCTCCAGACTTAAATGTATCACCACTAACCTGTAGGGG	3579						
QY	291	ACCCAATCTGACTCCTTCCCGCTTGGACATCGCAGCGCGGGAAGCAGTGC	350						
Db	3580	ACCCAATCTGACTCCTTCCCGCTTGGACATCGCAGCGCGGGAAGCAGTGC	3639						
QY	351	GGCTGGGCGCAGAGAGCTCCAGAGAGGCACTAGCGCTGTGGCGGAGGCTCGAC	410						
Db	3640	GGCTGGGCGCAGAGAGCTCCAGAGAGGCACTAGCGCTGTGGCGGAGGCTCGAC	3699						
QY	411	ATCCGACAGCAGCAGGAAAGTCTCTGGGCGATCTGTAATAAACCTTTTCTTT	470						
Db	3700	ATCCGACAGCAGCAGGAAAGTCTCTGGGCGATCTGTAATAAACCTTTTCTTT	3759						
QY	471	GTTTTAAAAA 482							
Db	3760	GTTTTAAAAA 3771							

RESULT 12
ABZ74587
ID ABZ74587 standard; DNA; 5033 BP.

XX	AC	ABZ74587;							
XX	DT	12-MAY-2003 (first entry)							
XX	DE	Secreted protein gene 360 genomic fragment HUFCL31, SEQ ID NO:1734.							
XX	XX	Human; secreted protein; cancer; tumour; hyperproliferative disorder;							
KW	KW	autoimmune disorder; inflammation; angiogenic diseases; AIDS;							
KW	KW	acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;							
KW	KW	drug screening; chromosome identification; chromosome mapping;							
KW	KW	cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;							
KW	KW	antianemic; vulnery; chromosome 9p21; gene; ds.							
XX	OS	Homo sapiens.							
XX	PN	WO200277013-A2.							
XX	PD	03-OCT-2002.							
XX	PF	26-MAR-2002; 2002WO-US009370.							
XX	PR	27-MAR-2001; 2001US-0278650P.							
PR	PR	12-SEP-2001; 2001US-00950082.							
PR	PR	12-SEP-2001; 2001US-00950083.							
XX	PA	(HUMA-) HUMAN GENOME SCI INC.							
XX	PI	Rosen CA, Ruben SM;							
XX	DR	WPI; 2003-040578/03.							
XX	PT	New human secreted proteins and nucleic acids, useful for detecting or							
PT	PT	treating cancer or other hyperproliferative disorders, autoimmune							
PT	PT	disorders, inflammatory disorders, HIV disease, hepatitis or anemia.							

XX PS Disclosure; Page 2315-2316; 2474pp; English.

XX CC ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
CC protein genes, and ABP0947-ABP01363 represent the proteins they encode.
CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
CC invention also encompasses antibodies specific for the secreted proteins,
CC the use of the secreted proteins in drug screening and recombinant
CC vectors and host cells comprising a nucleic acid of the invention. The
CC secreted proteins are thought to be involved in biological activities
CC associated with cellular signaling, cellular differentiation, cell
CC migration, prohormone activation and neurotransmitter activity. The
CC secreted proteins, nucleic acids encoding them, antibodies or antibody
CC fragments specific for the secreted proteins, and modulators of protein
CC activity are useful for diagnosing or treating cancers or other
CC hyperproliferative disorders. Additionally, the secreted proteins and
CC their nucleic acids may also be used in the treatment of autoimmune
CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
CC wound healing. Nucleic acids of the invention may be used for chromosome
CC identification, chromosome mapping, in gene therapy, for identifying
CC individuals from minute biological samples, as hybridisation probes, and
CC as molecular weight markers. The present sequence represents a human
CC secreted protein genomic fragment referred to in the disclosure of the
CC invention

XX SQ Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;

Query Match 80.0%; Score 398.4; DB 7; Length 5033;
Best Local Similarity 95.1%; Pred. No. 7.9e-101;
Matches 411; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY	51	ACTCAGTAGCAGCGCCGCTTCTGGGTCAACGCTGCAACCGCGCTCTGTCCTTCT	110						
Db	3340	ACTCCACATGCCCCCTCTTCTTTCATCCCGCCACAGGCGTCTGTGCTTCTCT	3399						
QY	111	CGAGGGGCGGTGTAGTCTCCAGTATGTTCCGCCACGCGCTCTTCCGACCTTCTGA	170						
Db	3400	CGAGGGGCGGTGTAGTCTCCAGTATGTTCCGCCACGCGCTCTTCCGACCTTCTGA	3459						
QY	171	CCAAAGCGCCAGGACTGCAGCCAGAGAGAGGGGCTCACTCTTATCTCGGACCC	230						
Db	3460	CCAAAGCGCCAGGACTGCAGCCAGAGAGAGGGGCTCACTCTTATCTCGGACCC	3519						
QY	231	ACTGCAAGCAGCGCGCTCTCCAGACTTAAATGTATCACCACTAACCTGTAGGGG	290						
Db	3520	ACTGCAAGCAGCGCGCTCTCCAGACTTAAATGTATCACCACTAACCTGTAGGGG	3579						
QY	291	ACCCAATCTGACTCCTTCCCGCTTGGACATCGCAGCGCGGGAAGCAGTGC	350						
Db	3580	ACCCAATCTGACTCCTTCCCGCTTGGACATCGCAGCGCGGGAAGCAGTGC	3639						
QY	351	GGCTGGGCGCAGAGAGCTCCAGAGAGGCACTGAGCGCTGTGGCGAGGCGCTCGAC	410						
Db	3640	GGCTGGGCGCAGAGAGCTCCAGAGAGGCACTGAGCGCTGTGGCGAGGCGCTCGAC	3699						
QY	411	ATCCGACAGCAGCAGGAAAGTCTCTGGGCGATCTGTAATAAACCTTTTCTTT	470						
Db	3700	ATCCGACAGCAGCAGGAAAGTCTCTGGGCGATCTGTAATAAACCTTTTCTTT	3759						
QY	471	GTTTTAAAAA 482							
Db	3760	GTTTTAAAAA 3771							

RESULT 13
ADC21005
ID ADC21005 standard; DNA; 5033 BP.

XX	AC	ADC21005;							
XX	DT	18-DEC-2003 (first entry)							

DE Human secreted protein-related DNA sequence #423.
XX
KW gene therapy; human; secreted protein; haemopoietic disorder;
KW haematological disorder; anaemia; haemophilia; inflammatory disorder;
KW inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
KW leukaemia; wound healing; epithelial cell proliferation disorder;
KW immune disorder; autoimmune disorder; asthmatic disorder;
KW cardiovascular disorder; atherosclerosis; myocarditis;
KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;
KW gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200292787-A2.
XX
PD 21-NOV-2002.
XX
PF 26-MAR-2002; 2002WO-US009257.
XX
PR 27-MAR-2001; 2001US-0278650P.
XX
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-129287/12.
XX
XX
PT New human secreted proteins and nucleic acid molecules, useful for
PT preparing a diagnostic or pharmaceutical composition for diagnosing,
PT preventing or treating hematopoietic or hematologic disorders, e.g.
PT anemia or hemophilia.
XX
PS Disclosure; SEQ ID NO 959; 1512pp; English.
XX
XX The invention comprises the amino acid and coding sequences of human
CC secreted proteins. The DNA and protein sequences of the invention are
CC useful for detecting, preventing, diagnosing, prognosticating, treating
CC or ameliorating: haematopoietic or haematological disorders (e.g. anaemia
CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease
CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
CC disorders (e.g. autoimmune disorders and asthmatic disorders);
CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);
CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
CC and gastrointestinal disorders (e.g. duodenal ulcers and
CC gastroenteritis). The present DNA sequence was used in the
CC exemplification of the invention.
XX
SQ Sequence 5033 BP; 1061 A; 1409 C; 1465 G; 1098 T; 0 U; 0 Other;

Query Match 80.0%; Score 398.4; DB 9; Length 5033;
Best Local Similarity 95.1%; Pred.No. 7.9e-101;
Matches 411; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 51 ACTCAGTAGAGCGCGCCTTCTGGGTCACGCTGGCAACCGGCTCTGTGCTCTTCT 110
DB 3340 ACTCCACATGCCCTCTTCTTTCGATCCCAACGCGACAGGCGTCTGTGCTCTTCT 3399
QY 111 CGAGAGGCGCGTGTAGTCTCCAGTATGTTGGCCAGCGCTCTTCCGACCTTCTGA 170
DB 3400 CGAGAGGCGCGTGTAGTCTCCAGTATGTTGGCCAGCGCTCTTCCGACCTTCTGA 3459
QY 171 CCAAGCGCCCAAGACTGTCAGCCAGAGAGAGAGGCGGCTCACCTTTATCTCGGCGACCC 230
DB 3460 CCAAGCGCCCAAGACTGTCAGCCAGAGAGAGAGGCGGCTCACCTTTATCTCGGCGACCC 3519
QY 231 ACTGCACAGCAGGCGCGCTCTCCAGACTTAATATGATACCACTAACCTGTGAGGGG 290
DB 3520 ACTGCACAGCAGGCGCGCTCTCCAGACTTAATATGATACCACTAACCTGTGAGGGG 3579
QY 291 ACCCAATCTGGAATCTTCCCGCGCTTGGGACATCGCAGGCGGGAAGCAGTGCCTGCCA 350

DB 3580 ACCCAATCTGGAATCTTCCCGCGCTTGGGACATCGCAGGCGGGAAGCAGTGCCTGCCA 3639
QY 351 GGCCTGGGCCAGAGAGAGCTCCAGAGAGGACACTGAGCGCTGTGGCGGAGGCTCGAC 410
DB 3640 GGCCTGGGCCAGAGAGAGCTCCAGAGAGGACACTGAGCGCTGTGGCGGAGGCTCGAC 3699
QY 411 ATCCGACAGCAGCAGGAAAGTCTCCTGGGGCGATCTGTAATAAACCTTTTCTTT 470
DB 3700 ATCCGACAGCAGCAGGAAAGTCTCCTGGGGCGATCTGTAATAAACCTTTTCTTT 3759
QY 471 GTTTTAAAAA 482
DB 3760 GTTTTAAAAA 3771

RESULT 14
AAH17543/C
ID AAH17543 standard; cDNA; 3180 BP.
XX
AC AAH17543;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:17021.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 17021; 2537pp + Sequence listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides, and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 3180 BP; 687 A; 881 C; 924 G; 688 T; 0 U; 0 Other;
Query Match 9.2%; Score 46; DB 4; Length 3180;
Best Local Similarity 57.7%; Pred. No. 0.022;
Matches 82; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 91 GCGCTCCTGCTCCTCTTCTCGAGGGGCGCTGTGAGTCTCCAGTATGTTCCGCCACG 150
DB 1564 GGGGTGTGATCTCGGAGTAGGGGAGACAGAGTTGGGCGCTCAGATGCTGGCACAGC 1505
QY 151 GCTCTTCGACCCCTTCTGACCAAGGCCAAGGACTGCAGCCAGAGAGAGGGGGCTCA 210
DB 1504 AGGGTCCCTACCCCTGTGGGCCCAATGTGATAGCACTGAGGCCGAGATAGGGGGCTTT 1445
QY 211 CCTCTTATCTCGGGCAGCCAC 232
DB 1444 AGAGTAGGCCGCTGGAGCCTC 1423
RESULT 15
AAL03901/c
ID AAL03901 standard; DNA; 5049 BP.
XX
AC AAL03901;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 6589.
XX
KM Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226682P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.

Search completed: February 23, 2004, 19:50:07
Job time : 204.357 secs

PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250391P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen is
XX
PT used in preventing, treating or ameliorating a medical condition.

XX
XX
PS Disclosure; SEQ ID NO 6589; 1297pp + Sequence Listing; English.
XX
XX
CC The present invention provides the protein and coding sequences of a
XX
CC number of human reproductive system related antigens. These can be used
XX
CC in the prevention and treatment of reproductive system disorders,
XX
CC including cancer. The present sequence is a genomic sequence encoding a
XX
XX protein of the invention

SQ Sequence 5049 BP; 1248 A; 1315 C; 1438 G; 1048 T; 0 U; 0 Other;

Query Match 9.2%; Score 46; DB 4; Length 5049;
Best Local Similarity 57.7%; Pred.No. 0.027;
Matches 82; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 91 GGGCTCTGTCGCTTCTTCCTCGAGGGCCGCTGTGAGTCTCCAGTATGTTGGCCAGC 150
DB 1575 GGGGTGTTGATCCTGGAGTAGGGGAGACAGAGTTGGGCTTCAGATGCTGGCAGC 1516
QY 151 GCTCTTCGACCCCTTTCGACCAAGCCCAAGAGAGTGCAGCCAGAGAGAGGGGCTCA 210
DB 1515 AGGTCCTCACCCCTGTGGGCCCCAATGTGATAAGCACTGAGGCCGAGATAGGGGCTTT 1456
QY 211 CCTCTATCCTCGGCGACCCAC 232
DB 1455 AGAGTAGGCCGCTGGGAGCCTC 1434

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 19:35:58 ; Search time 45.2027 Seconds
(without alignments)
6113.919 Million cell updates/sec

Title: US-09-936-456-3

Perfect score: 498

Sequence: 1 cctcagcgtgcgcgtctgcc.....aaaaaataaagtcgacc 498

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/2/ina/PCtUS_COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36.6	7.3	753	US-09-252-991A-93	Sequence 93, Appl
2	36.6	7.3	846	US-09-252-991A-102	Sequence 102, App
3	36.6	7.3	1038	US-09-252-991A-98	Sequence 98, Appl
4	36.6	7.3	1398	US-09-252-991A-91	Sequence 91, Appl
5	36.6	7.3	2934	US-09-252-991A-84	Sequence 84, Appl
6	36	7.2	504	US-09-252-991A-3564	Sequence 3564, Ap
7	36	7.2	1002	US-09-252-991A-3539	Sequence 3539, Ap
8	36	7.2	1644	US-09-252-991A-3548	Sequence 3548, Ap
9	36	7.2	2832	US-09-252-991A-3523	Sequence 3523, Ap
10	35.4	7.1	408	US-09-252-991A-11606	Sequence 11606, A
11	35.4	7.1	564	US-09-252-991A-11689	Sequence 11689, A
12	35.4	7.1	1080	US-09-252-991A-11765	Sequence 11765, A
13	35.4	7.1	2080	US-09-252-991A-11765	Sequence 11765, A
14	35	7.0	810	US-09-252-991A-15239	Sequence 15239, A
15	35	7.0	915	US-09-252-991A-15234	Sequence 15234, A
16	35	7.0	1269	US-09-252-991A-15226	Sequence 15226, A
17	34.2	6.9	444	US-09-252-991A-4073	Sequence 4073, Ap
18	34.2	6.9	762	US-09-252-991A-4088	Sequence 4088, Ap
19	34.2	6.9	1305	US-09-252-991A-4057	Sequence 4057, Ap
20	34.2	6.9	1347	US-09-252-991A-4100	Sequence 4100, Ap
21	34	6.8	1382	US-09-252-991A-4100	Sequence 4100, Ap
22	34	6.8	2493	US-09-252-991A-11987	Sequence 11987, A
23	34	6.8	2991	US-09-252-991A-12025	Sequence 12025, A
24	33.6	6.7	2159	US-08-286-870A-7	Sequence 7, Appli
25	33.4	6.7	7218	US-08-222-463-14	Sequence 14, Appl
26	33.2	6.7	154	US-09-621-976-11503	Sequence 11503, A
27	33.2	6.7	432	US-09-252-991A-3530	Sequence 3530, Ap

28	33	6.6	834	4	US-09-621-976-2574	Sequence 2574, Ap
29	33	6.6	441529	3	US-09-103-840A-1	Sequence 1, Appli
30	32.8	6.6	540	4	US-09-252-991A-11862	Sequence 11862, A
31	32.8	6.6	1188	4	US-09-252-991A-11800	Sequence 11800, A
32	32.6	6.5	2290	4	US-09-149-476-177	Sequence 177, App
33	32.6	6.5	49272	1	US-08-614-770A-1	Sequence 1, Appli
34	32.6	6.5	4403765	3	US-09-103-840A-2	Sequence 2, Appli
35	32.6	6.5	441529	3	US-09-103-840A-1	Sequence 1, Appli
36	32.4	6.5	663	4	US-09-252-991A-9138	Sequence 9138, Ap
37	32.4	6.5	1056	4	US-09-252-991A-8728	Sequence 8728, Ap
38	32.2	6.5	597	4	US-09-613-303-34	Sequence 34, Appl
39	32.2	6.5	597	4	US-10-267-311-34	Sequence 34, Appl
40	32.2	6.5	2468	1	US-08-333-358-11	Sequence 11, Appl
41	32.2	6.5	2468	1	US-08-463-694-11	Sequence 11, Appl
42	32.2	6.5	2468	1	US-08-694-501-11	Sequence 11, Appl
43	32	6.4	4403765	3	US-09-103-840A-2	Sequence 2, Appli
44	31.8	6.4	224	4	US-09-016-434-212	Sequence 212, App
45	31.8	6.4	1053	4	US-09-252-991A-14223	Sequence 14223, A

ALIGNMENTS

```
RESULT 1
US-09-252-991A-93
Sequence 93, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 93
LENGTH: 753
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-93

Query Match          7.3%; Score 36.6; DB 4; Length 753;
Best Local Similarity 60.6%; Pred. No. 0.65;
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 318 GGCACATGCGAGCCGGGAGAGCAGTCCCGCCAGGCTGGCCAGAGAGCTCCAGAG 377
Db 285 GCGCCGTGCGGACCCGACGCGCTTCGCTCGGCGCCAGAGGCTGTGCCGAGTAC 344

QY 378 GGCACGTAGCGCTGCTGGCGCGAGGCGCTCGACATCCGC 416
Db 345 AGCTGACGCCCGACCTTGCGCAAGCGCTGGCCACCGGC 383

RESULT 2
US-09-252-991A-102
Sequence 102, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
```

; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 102
; LENGTH: 846
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-102

Query Match 7.3%; Score 36.6; DB 4; Length 846;
Best Local Similarity 60.6%; Pred. No. 0.68;
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 318 GGGACATCGCAGCGCCGGGAAGCAGTGGCCCGCCAGGCTGGGCCAGAGAGCTCCAGGAAG 377
Db 727 GCGCCGTCGCGGACCCGACGCGCTTCTCGCTGGGGCCAGCAGGCTGTGCTCCGAGGTAC 786
QY 378 GGCACGTGAGCGCTGCTGGCGCGAGGCTTCGACATCCGC 416
Db 787 AGCTCGACGCCCGACGCTTGCGCAAGGCTTGCGCCACCGGC 825

RESULT 3
US-09-252-991A-98
; Sequence 98, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 98
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-98

Query Match 7.3%; Score 36.6; DB 4; Length 1038;
Best Local Similarity 60.6%; Pred. No. 0.74;
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 318 GGGACATCGCAGCGCCGGGAAGCAGTGGCCCGCCAGGCTGGGCCAGAGAGCTCCAGGAAG 377
Db 737 GCGCGTCGCGGACCCGACGCGCTTCTCGCTGGGGCCAGCAGGCTGTGCTCCGAGGTAC 796
QY 378 GGCACGTGAGCGCTGCTGGCGCGAGGCTTCGACATCCGC 416
Db 797 AGCTCGACGCCCGACGCTTGCGCAAGGCTTGCGCCACCGGC 835

RESULT 4
US-09-252-991A-91/c
; Sequence 91, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 91
; LENGTH: 1398

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-91

Query Match 7.3%; Score 36.6; DB 4; Length 1398;
Best Local Similarity 60.6%; Pred. No. 0.83;
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 318 GGGACATCGCAGCGCCGGGAAGCAGTGGCCCGCCAGGCTGGGCCAGAGAGCTCCAGGAAG 377
Db 767 GCGCCGTCGCGGACCCGACGCGCTTCTCGCTGGGGCCAGCAGGCTGTGCTCCGAGGTAC 708
QY 378 GGCACGTGAGCGCTGCTGGCGCGAGGCTTCGACATCCGC 416
Db 707 AGCTCGACGCCCGACGCTTGCGCAAGGCTTGCGCCACCGGC 669

RESULT 5
US-09-252-991A-84/c
; Sequence 84, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 84
; LENGTH: 2934
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-84

Query Match 7.3%; Score 36.6; DB 4; Length 2934;
Best Local Similarity 60.6%; Pred. No. 1.1;
Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 318 GGGACATCGCAGCGCCGGGAAGCAGTGGCCCGCCAGGCTGGGCCAGAGAGCTCCAGGAAG 377
Db 2055 GCGCGTCGCGGACCCGACGCGCTTCTCGCTGGGGCCAGCAGGCTGTGCTCCGAGGTAC 1996
QY 378 GGCACGTGAGCGCTGCTGGCGCGAGGCTTCGACATCCGC 416
Db 1995 AGCTCGACGCCCGACGCTTGCGCAAGGCTTGCGCCACCGGC 1957

RESULT 6
US-09-252-991A-3564/c
; Sequence 3564, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3564
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3564

```
Query Match      7.2%; Score 36; DB 4; Length 504;
Best Local Similarity 51.9%; Pred. No. 0.81;
Matches 81; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
```

QY	287	GGGGACCCAAATCTG	ACTTCTTCCCCCGCTTGGACATCGCAGGCCGGGAACGATGCC	346
Db	193	GGGGAGCGGGGCATG	AGTCTCGACCGAGCGGGCGGCGCGCCCTTGGGGGTGGCGCC	134
QY	347	GCCAGGCTGGGCCAGAGAGCTCCAGGAAGGCACTGAGCGCTGCTGCGCGAGGCTTC	406	
Db	133	ACGCTGACAGTCCCGTGCTGCGCCACAGGCGCGCTGGAGGCGCTGTGCTGCATCCACTTC	74	
QY	407	GGACATCCGCGAGGACACAGGAAAGTCTCCTGGGCG	442	
Db	73	GCCCATCCGAGGCGCTCGGCGAAGATTGCGCGCG	38	

RESULT 7

US-09-252-991A-3539
; Sequence 3539, Application US/09252991A

```

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3539
; LENGTH: 1002
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3539

```

```

Query Match      7.2%; Score 36; DB 4; Length 1002;
Best Local Similarity 51.9%; Pred. No. 1.1;
Matches 81; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```

QY 287 GGGGACCCAACTGTGACTTCCTTCCCCCGCTTTGGCATCGCAGGCCGGGAACAGTGCCC 346
| | | | |
Db 281 GGGGAGCGCGGCATGAGTCTCGACGAGSGCGGCGAGCCCGCGCTGGGGCTGGGGCGCC 340
| | | | |
QY 347 GCCAGGCTGGGCCAGGAGAGCTCCAAGAAGGCACTGAGCGTCTGTCGGCGAGGCGTTC 406
| | | | |
Db 341 ACGCTGCAAGTCCCGGTGCTGCGCACAGGCCGCGTGGAGGCGCTGTGCGATTCCACTTC 400
| | | | |
QY 407 GGACATCCGACAGGCACCAAGAAAGTCTCTGCGGC 442
| | | | |
Db 401 GCCCATCCGACGGCGCTCGGGCGAAGATTGCGCGCG 436

RESULT 8

US-09-252-991A-3548/c
; Sequence 3548, Application US/09252991A
; Patent No. 6551795

```

; GENERATION INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

```

```
; SEQ ID NO 3548
;
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
;
US-09-252-991A-3548
```

Query Match	7.2%;	Score 36;	DB 4;	Length 1644;
Best Local Similarity	51.9%;	Pred. No. 1.3;		
Matches	81;	Conservative	0;	Mismatches 75;
				Indels 0;
				Gaps 0;

QY	287	GGGGACCCCAATCTGGA	CTCTTCCCTCCGCTTTGG	GCATCGCAGGCCGGGA	ACAGTAGTCCC	346
Db	816	GGGGAGGGCGGCATG	AGTCTCGACGAGGGCG	CGACCCGCGCGCTGG	GGCGCC	757
QY	347	GCCAGGCTGGGCCAG	AGAGCTCCAGGAAGG	CACTGAGCGCTGTG	CGCGAGGCTC	406
Db	756	ACGCTGCACGTCCCG	GTGCTGCGCACGGCG	CGACTGAGGCGCTG	CTGCGGATCCACTTC	697
QY	407	GGACATCCGACAGG	CAACGAGGAAGTCT	CTCGGGC	442	
Db	696	GCCCATCCGACAGG	CGCTCGCGAAGATT	CGCGCGC	661	

RESULT 9

US-09-252-991A-3523
; Sequence 3523, Application US/09252991A

```

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3523
; LENGTH: 2832
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3523

```

Query Match	7.2%	Score 36;	DB 4;	Length 2832;
Best Local Similarity	51.9%	Pred. No. 1.6;		
Matches 81; Conservative	0;	Mismatches 75;	Indels 0;	Gaps 0

QY 287 GGGGACCCCAATCTGGACTCCTTCCCCGCCCTTTGGGACATTCGAGGCCGGGAAGCAATGCC 346
| | | | |
Db 1054 GGGGAGCGCGGCATGAGTCTGCACGAGGCGCGGAGCGCGCGCGCTTGGGGCTGGGGGCC 1133
| | | | |
QY 347 GCCAGCGCTGGGCCAAGAGAAGCTCCAGGAAGGGCACTGAGCGCTGTGGCGGAGGCGTC 406
| | | | |
Db 1114 ACGTGACAGTCCC GG TGTGGGCCACGCGCGGCTGGAAGCGCTGTGGCATCACTTC 1173
| | | | |
QY 407 GGACATCCGACAGCACGAGGAAGTCTCCTGGGGC 442
| | | | |
Db 1174 GCCCATCCGACAGCGCTCGGCGAAGATTGCGCGCGC 1209

RESULT 10

US-09-252-991A-11606
; Sequence 11606, Application US/09252991A

;
 ; PATENT NO. 6551/95
 ;
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT: Marc J. Rubenfield et al.
 ;
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ;
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ;
 ; FILE REFERENCE: 107196.136
 ;
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ;
 ; CURRENT FILING DATE: 1999-02-18
 ;

;; PRIOR APPLICATION NUMBER: US 60/074,788
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 11606
;; LENGTH: 408
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11606

Query Match 7.1%; Score 35.4; DB 4; Length 408;
Best Local Similarity 59.4%; Pred. No. 1.1;
Matches 60; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 20 CCGCTCCCGCCTAGGCTCCTCCGGGCTCACCACCTAGTACGAGCGCGCCTTCTGGGTCA 79
Db 67 CCG 126
QY 80 CGCTGGCAACCGCGCTCTGTGCTCTTCTCTCGAGGGGCC 120
Db 127 CGCTGGCAACCGCGCTCTGTGCTCTTCTCTCGAGGGGCC 167

RESULT 11
US-09-252-991A-11689/c
;; Sequence 11689, Application US/09252991A
;; Patent No. 6551795
;; GENERAL INFORMATION:
;; APPLICANT: Marc J. Rubenfield et al.
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;; FILE REFERENCE: 107196.136
;; CURRENT APPLICATION NUMBER: US/09/252,991A
;; PRIOR FILING DATE: 1999-02-18
;; PRIOR APPLICATION NUMBER: US 60/074,788
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 11689
;; LENGTH: 564
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11689

Query Match 7.1%; Score 35.4; DB 4; Length 564;
Best Local Similarity 59.4%; Pred. No. 1.3;
Matches 60; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 20 CCGCTCCCGCCTAGGCTCCTCCGGGCTCACCACCTAGTACGAGCGCGCCTTCTGGGTCA 79
Db 161 CCG 102
QY 80 CGCTGGCAACCGCGCTCTGTGCTCTTCTCTCGAGGGGCC 120
Db 101 CGCTGGCAACCGCGCTCTGTGCTCTTCTCTCGAGGGGCC 61

RESULT 12
US-09-252-991A-11765/c
;; Sequence 11765, Application US/09252991A
;; Patent No. 6551795
;; GENERAL INFORMATION:
;; APPLICANT: Marc J. Rubenfield et al.
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;; FILE REFERENCE: 107196.136
;; CURRENT APPLICATION NUMBER: US/09/252,991A
;; PRIOR FILING DATE: 1999-02-18
;; PRIOR APPLICATION NUMBER: US 60/074,788
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190

;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 11765
;; LENGTH: 1080
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11765

Query Match 7.1%; Score 35.4; DB 4; Length 1080;
Best Local Similarity 59.4%; Pred. No. 1.6;
Matches 60; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 20 CCGCTCCCGCCTAGGCTCCTCCGGGCTCACCACCTAGTACGAGCGCGCCTTCTGGGTCA 79
Db 373 CCG 314
QY 80 CGCTGGCAACCGCGCTCTGTGCTCTTCTCTCGAGGGGCC 120
Db 313 CGCTGGCAACCGCGCTCTGTGCTCTTCTCTCGAGGGGCC 273

RESULT 13
US-09-716-129-17
;; Sequence 17, Application US/09716129
;; Patent No. 6632920
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: 36 Human Secreted Proteins
;; FILE REFERENCE: P2025P1
;; CURRENT APPLICATION NUMBER: US/09/716,129
;; PRIOR FILING DATE: 2000-11-17
;; PRIOR APPLICATION NUMBER: 60/076,053
;; PRIOR FILING DATE: 1998-02-26
;; PRIOR APPLICATION NUMBER: 60/076,057
;; PRIOR FILING DATE: 1998-02-26
;; PRIOR APPLICATION NUMBER: 60/076,052
;; PRIOR FILING DATE: 1998-02-26
;; PRIOR APPLICATION NUMBER: 60/076,054
;; PRIOR FILING DATE: 1998-02-26
;; PRIOR APPLICATION NUMBER: 60/076,051
;; PRIOR FILING DATE: 1998-02-26
;; NUMBER OF SEQ ID NOS: 186
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 17
;; LENGTH: 2080
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: SITE
;; LOCATION: (145)
;; OTHER INFORMATION: n equals a,t,g, or c
US-09-716-129-17

Query Match 7.1%; Score 35.4; DB 4; Length 2080;
Best Local Similarity 52.8%; Pred. No. 2.1;
Matches 75; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 18 GCCCGCTCCCGCCTAGGCTCCTCCGGGCTCACCACCTAGTACGAGCGCGCCTTCTGGGT 77
Db 45 GCGCTCTCTCCCTTCCCTCCCTCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104
QY 78 CACGCTGGCAACCGCGCTCTGTGCTCTTCTCTCGAGGGGCCGTGTGAGTCTCCAGTA 137
Db 105 CGAAGCGAAGATGGCACTTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 164
QY 138 TGTTCGGCGCGCGCTCTTCTCTCGAGGGGCCGTGTGAGTCTCCAGTA 159
Db 165 TATTGGCTCTTACTACTGGC 186

RESULT 14
US-09-252-991A-15239
;; Sequence 15239, Application US/09252991A

Search completed: February 23, 2004, 23:23:33
Job time : 55.2027 secs

```

; Patent NO. 6551795
;
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15239
; LENGTH: 810
;
; TYPE: DNA
;
; ORGANISM: Pseudomonas aeruginosa
;
US-09-252-991A-15239

```

Query Match	7.0%;	Score 35;	DB 4;	Length 810;
Best Local Similarity	54.2%;	Pred. No. 1.9;		
Matches 71; Conservative	0;	Mismatches 60;	Indels 0;	Gaps 0;

QY	318	GGGACATCGAGAGCCCGGGAAAGCAAGTCCCCCGCAGCCTGGGCCAGAGAGCTTCCAGGAAG	377
Db	176	GGGTATTCCGCCGAGCGCAACTGAGAGCCCGGCGACACTGGCTGGGGCTCGGCCCCGGCCG	235
QY	378	GGCACTGAGCGGTCTGTGGCGCGAGGCTTCGACATCCGACAGGACCAAGGAAGTCTCT	437
Db	236	GCCGTTTCGCCGCCCTGACCAAGTCGCGATCCGTCGACAGGCTTCTCGCAAGCGCTCC	295
QY	438	GGGGCGATCTG	448
Db	296	GTGGCGAAGCTG	306

```

RESULT 15
US-09-252-991A-15234
; Sequence 15234, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15234
; LENGTH: 915
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15234

```

	Query Match	7.0%;	Score 35;	DB 4;	Length 915;
	Best Local Similarity	54.2%;	Pred. No. 2;		
	Matches	71; Conservative	0; Mismatches	60; Indels	0; Gaps
QY	318	GGGACATGCAGCGCCGGGAAGCAGTGTCCCCGCCAGGCTTGSGCCAGAGAGCTCCAGGAAG	377		
Db	54	GGGTATTGCCCGACCGCAGACTGAGGCGCGCGCACCTGGCTTGGGCTTCGGCCGCGCG	113		
QY	378	GGCACTGAGCGCTGTGTGGCGGAGGCTTCGACATCCGCAGGACACCAGGAAAGTCTCCT	437		
Db	114	GCCGTTTCGCGCGCCCTGACCAACGTCGCGATTCGTGCGAGGCTCTCGGCAAGCGCTCCC	173		
QY	438	GGGGCGATCTG	448		
Db	174	GTTGGCGAAGTGT	184		

QY	318	GGGACATCGCAGGCCCGGAA	GCAGTCCCTCCCGCAGAGAGCTCCAGGAAG	377
Db	54	GGGTATTCCGCCGACGCGAC	CTGGAGGCGCGCGGCACTGGCTGGGGCTCCGCCCGGCG	113
QY	378	GGCACTGAGCGCTGCTGCGCG	GAGGCTTCGACATCCGACAGGCA	CCAGGAAAGTCTCCT
Db	114	GCCGTTTCGCGCCCTGACCA	AGTCCGCGATCCGTCGACAGGCT	CTCGGCAAGGCTCCC
QY	438	GGGCGGATCTG	448	
Db	174	GTGGCGAACTG	184	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 19:36:48 ; Search time 191.084 Seconds
(without alignments)
9125.925 Million cell updates/sec

Title: US-09-936-456-3
Perfect score: 498
Sequence: 1 ctctagcgctgcgcgtctgccc.....aaaaaaataaagtcgacc 498

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2308684 seqs, 1750822206 residues
Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	465.4	93.5	1474	US-10-264-237-1097	Sequence 1097, Ap
2	453	91.0	1421	US-10-187-657-2	Sequence 2, Appli
3	447.8	89.9	1420	US-10-187-657-4	Sequence 4, Appli
4	398.4	80.0	2684	US-10-094-749-1195	Sequence 1195, Ap
5	266	53.4	346	US-10-187-657-3	Sequence 3, Appli
6	158.2	31.8	1594	US-10-187-657-9	Sequence 9, Appli
7	65.2	13.1	439	US-10-027-632-91970	Sequence 91970, A
8	49.8	10.0	594	US-10-123-155-10	Sequence 10, Appli
9	49.8	10.0	594	US-10-146-731-10	Sequence 10, Appli
10	49.8	10.0	594	US-10-140-472-10	Sequence 10, Appli
11	49.8	10.0	594	US-10-141-761-10	Sequence 10, Appli
12	49.8	10.0	594	US-10-142-885-10	Sequence 10, Appli
13	49.8	10.0	594	US-10-158-790-10	Sequence 10, Appli
14	49.8	10.0	594	US-10-137-871-10	Sequence 10, Appli
15	49.8	10.0	594	US-10-140-923-10	Sequence 10, Appli

C	16	49.8	10.0	594	15	US-10-141-756-10	Sequence 10, Appli
C	17	49.8	10.0	594	15	US-10-141-759-10	Sequence 10, Appli
C	18	49.8	10.0	594	15	US-10-140-805-10	Sequence 10, Appli
C	19	49.8	10.0	594	15	US-10-140-864-10	Sequence 10, Appli
C	20	48.8	9.8	450	14	US-10-187-657-7	Sequence 7, Appli
C	21	46	9.2	5049	10	US-09-764-891-6588	Sequence 6588, Ap
C	22	46	9.2	5049	10	US-09-764-891-6589	Sequence 6589, Ap
C	23	46	9.2	5049	14	US-10-091-572-460	Sequence 460, App
C	24	46	9.2	5049	14	US-10-091-572-461	Sequence 461, App
C	25	42	8.4	536	16	US-10-338-110-119	Sequence 119, App
C	26	41.6	8.4	1341	14	US-10-156-761-7235	Sequence 7235, Ap
C	27	41.6	8.4	9025608	14	US-10-156-761-1	Sequence 1, Appli
C	28	39	7.8	536	16	US-10-338-110-119	Sequence 119, App
C	29	38.8	7.8	620	15	US-10-027-632-197718	Sequence 197718, Appli
C	30	38.6	7.8	9025608	14	US-10-156-761-1	Sequence 1, Appli
C	31	38.2	7.7	671	14	US-10-184-644-346	Sequence 346, App
C	32	38.2	7.7	671	14	US-10-184-644-346	Sequence 346, App
C	33	36.4	7.3	431	12	US-10-424-599-20038	Sequence 20038, A
C	34	36.4	7.3	1751	15	US-10-264-237-954	Sequence 954, App
C	35	36.2	7.3	535	15	US-10-027-632-51283	Sequence 51283, A
C	36	36.2	7.3	732	14	US-10-156-761-1457	Sequence 6733, Ap
C	37	36.2	7.3	2776	10	US-09-764-891-6731	Sequence 6731, Ap
C	38	36.2	7.3	2776	10	US-09-764-891-6733	Sequence 6733, Ap
C	39	35.8	7.2	795	15	US-10-260-238-5695	Sequence 5695, Ap
C	40	35.6	7.1	855	15	US-10-402-842-46	Sequence 46, Appli
C	41	35.6	7.1	47988	15	US-10-402-842-1	Sequence 1, Appli
C	42	35.4	7.1	63158	14	US-10-292-198-1	Sequence 1, Appli
C	43	35.2	7.1	879	14	US-10-156-761-2933	Sequence 2933, Ap
C	44	35.2	7.1	1116	15	US-10-369-493-41648	Sequence 41648, A
C	45	35.2	7.1	1395	14	US-10-205-032-21	Sequence 21, Appli

ALIGNMENTS

RESULT 1
US-10-264-237-1097
; Sequence 1097, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131PI
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: US/10/264,237
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 1097
; LENGTH: 1474
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-264-237-1097

QY	1	CTCTAGCGTGGCGCTCTGCGGCTCCGCGCTCTGCGGCTCAGTACG	60
DB	965	CTCTAGCGTGGCGCTCTGCGGCTCCGCGCTCTGCGGCTCAGTACG	1023
QY	61	AGCGCGCTCTGCGGCTCAGCGTGGCAACCGGCGTCTGCGGCTCTGCGGCGG	120
DB	1024	-GCGCGCGCTCTGCGGCTCAGCGTGGCAACCGGCGTCTGCGGCGG	1082
QY	121	GTGGTAGTCTCCAGTATGTTGGGCGGCGGCTCTGCGGCTCTGCGGCGG	180
DB	1083	GTGGTAGTCTCCAGTATGTTGGGCGGCGGCTCTGCGGCTCTGCGGCGG	1142

QY	181	AAGAGCTGCAGCCAGGAGAGAGGGGGCTCACTCTTATCCTCGGCGAACCACTGCACAAAG	240
Db	1143	AAAGACTGCAGCCAGGAGAGAGGGGGCTCACTCTTATCCTCGGCGAACCACTGCACAAAG	1202
QY	241	CAGGCCGCTCTCCAGACTTAATAATGTATCACTAACCTGTGAGGGGGACCAATCTG	300
Db	1203	CAGGCCGCTCTCCAGACTTAATAATGTATCACTAACCTGTGAGGGGGACCAATCTG	1262
QY	301	GACTCCTTCCCGCCTTGGAACATCCGAGGCGGGGAAGCATGTGCCCGCCAGGCTTGAGCC	360
Db	1263	GACTCCTTCCCGCCTTGGAACATCCGAGGCGGGGAAGCATGTGCCCGCCAGGCTTGAGCC	1322
QY	361	AGGAGAGCTCCAGGAAGGCACTGAGCGTGTGGCGGAGGCTCGACAATCCGAGGC	420
Db	1323	AGGAGAGCTCCAGGAAGGCACTGAGCGTGTGGCGGAGGCTCGACAATCCGAGGC	1382
QY	421	ACCAGGGAAAGTCTCCTGGGCGAATCTGTAATAAACCTTTTTTCTTTGTTTTTAA	480
Db	1383	ACCAGGGAAAGTCTCCTGGGCGAATCTGTAATAAACCTTTTTTCTTTGTTTTTAA	1442
QY	481	AAAAAAAAATAAA 491	
Db	1443	AAAAAAAAAAAAA 1453	

RESULT 2

```

US-10-187-657-2
/ Sequence 2, Application US/10187657
/ Publication No. US20030068311A1
/
GENERAL INFORMATION:
/ APPLICANT: Lasek, Amy K. W.
/ APPLICANT: Baughn, Mariah R.
/ APPLICANT: Azimzai, Yalda
/ TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
/ FILE REFERENCE: PV-0009 CIP
/ CURRENT APPLICATION NUMBER: US/10/187,657
/ CURRENT FILING DATE: 2002-07-01
/ PRIOR APPLICATION NUMBER: PCT/US00/07817
/ PRIOR FILING DATE: 2000-03-22
/ PRIOR APPLICATION NUMBER: 60/139,565
/ PRIOR FILING DATE: 1999-06-16
/ NUMBER OF SEQ ID NOS: 9
/ SOFTWARE: PERL Program
/ SEQ ID NO 2
/ LENGTH: 1421
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No. US20030068311A1 4901066CB1
US-10-187-657-2

```

Query Match	91.0%;	Score 453;	DB 14;	Length 1421;
Best Local Similarity	99.6%;	Pred. No. 1.1e-126;		
Matches 475;	Conservative 0;	Mismatches 0;	Indels 2;	Gaps 2;

QY	1	CTCTAGCGGTGCGGCTCTGCGCGCTCCGCGCTAGGCTCTCCGCGCTCACCACCTCAGTACG	60
Db	947	CTCTAGCGGTGCGGCTCTGCGCGCT-CCGCGTAGGCTCTCCGCGCTCACCACCTCAGTACG	1005
QY	61	AGCGCGGCTTCTGGGTACGCTGGCAACCGCGCTGTGCTCTTCTCTGGAGGGGCC	120
Db	1006	-GCGCGGCTTCTGGGTACGCTGGCAACCGCGCTGTGCTCTTCTCTGGAGGGGCC	1064
QY	121	GTGTGAGTCTTCCAGTATGTTCCGGCCACGCGCTCTTGCACACCTTCTGACCAAGCGCC	180
Db	1065	GTGTGAGTCTTCCAGTATGTTCCGGCCACGCGCTCTTGCACACCTTCTGACCAAGCGCC	1124
QY	181	AAGGACTGCAGCCAGGAGAGAGGGGCTCACCTCTTATCCTCGGCGACCCCACTGCACAAAG	240
Db	1125	AAGGACTGCAGCCAGGAGAGAGGGGCTCACCTCTTATCCTCGGCGACCCCACTGCACAAAG	1184
QY	241	CAGGCGGCTTCTCCCACTTTAAATGTATCACCACTAACTGTGAGGGGGACCACTCTG	300

Db	1185	CAGGCGCTCTCCAGACTTAAATGTATACCACTAACTGTGAGGGGACCCAAATCTG	1244
QY	301	GACTCCTTCCCGCCTTGGGACATCGCAGCGCCGGGAAGAGTGCCCGCCAGGCTTGGGCC	360
Db	1245	GACTCCTTCCCGCCTTGGGACATCGCAGCGCCGGGAAGAGTGCCCGCCAGGCTTGGGCC	1304
QY	361	AGGAGAGCTCCAGGAAGGGCACTGAGCGCTGCTGGCGCAGAGGCTCGGAATCCGAGGC	420
Db	1305	AGGAGAGCTCCAGGAAGGGCACTGAGCGCTGCTGGCGCAGAGGCTCGGAATCCGAGGC	1364
QY	421	ACCAGGGAAGTCTCCTGGGGCGATCTGTAAATAAACCTTTTTCTTTGTTTTTT	477
Db	1365	ACCAGGGAAGTCTCCTGGGGCGATCTGTAAATAAACCTTTTTCTTTGTTTTTT	1421

RESULT 3

```

US-10-187-657-4
; Sequence 4, Application US/10187657
; Publication No. US20030068311A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzal, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187,657
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139,565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
;
; SEQ ID NO 4
;
; LENGTH: 1420
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
;
; FEATURE:
;
; NAME/KEY: misc feature
;
; OTHER INFORMATION: Incyte ID No. US20030068311A1 3221661CA2
;
US-10-187-657-4

```

Query Match	89.9%	Score 447.8;	DB 14;	Length 1420;
Best Local Similarity	99.2%;	Pred. No. 4.2e-125;		
Matches 471;	Conservative 0;	Mismatches 2;	Indels 2;	Gaps 2;

QY	1	CTCTAGCGGTGCGGCTCTGCGCGCTCCGCGCTTAGGCTCCTCGCGCTCACCACTCAGTACG	60
Db	947	CTCTAGCGGTGCGGCTCTGCGCGCT - CCGCTTAGGCTCCTCGCGCTCACCACTCAGTACG	1005
QY	61	AGCGCGGCTTCTGGGTCA CGGTGGCAACCGGCGTCTGTGCTCTTCTCGAGGGGCC	120
Db	1006	-GCGCGGCTTCTGGGTCA CGGTGGCAACCGGCGTCTGTGCTCTTCTCGAGGGGCC	1064
QY	121	GTGTGAGTCTCCAGTATGTTGCGGCCAGCGCTCTCGACCTTCTGACCAAGCGCC	180
Db	1065	GTGTGAGTCTCCAGTATGTTGCGGCCAGCGCTCTCGACCTTCTGACCAAGCGCC	1124
QY	181	AAGCACTGCAGCCAGAGAGAGAGGGGCTCACTCTTATCTCGGCGAACCCACTGCACAAG	240
Db	1125	AAGCACTGCAGCCAGAGAGAGGGGCTCACTCTTATCTCGGCGAACCCACTGCACAAG	1184
QY	241	CAGGCGGCTCTCCAGACTTAAATGTATCACCACTAACCTGTGAGGGGGAACCAATCTG	300
Db	1185	CAGGCGGCTCTCCAGACTTAAATGTATCACCACTAACCTGTGAGGGGGAACCAATCTG	1244
QY	301	GACTCTTCCCCGCTTGGGACATCGCAGGCGCGGGAAGCAGTGCCTCCGCAAGGCTGGGCC	360
Db	1245	GACTCTTCCCCGCTTGGGACATCGCAGGCGCGGGAAGCAGTGCCTCCGCAAGGCTGGGCC	1304
QY	361	AGGAGAGCTCCAGGAAGGCACTGAGCGCTGTGGCGGAGGCTTCGACATTCGCAAGGC	420

Db 1305 AGGAGAGCTCCAGGAAGGCGACTGAGCGCTGCTGGCGGAGGCGCTCGACATCCGACAGGC 1364

QY 421 ACCAGGGAAGTCTCCTGGGGCGATCTGTAAATAAACCTTTTCTTTCTTTT 475

Db 1365 ACCAGGGAAGTCTCCTGGGGCGATCTGTAAATAAACCTTTTCTTTCTTTT 1419

RESULT 4

US-10-094-749-1195/c

; Sequence 1195, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094, 749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350, 435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1195
; LENGTH: 2684
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-1195

Query Match 80.0%; Score 398.4; DB 15; Length 2684;
Best Local Similarity 95.1%; Pred. No. 5.1e-110;
Matches 411; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 51 ACTCAGTACGAGCGCGCCTTCTGGGTACAGCTGGCAACCGCGCTCTGTGCTTCTCT 110
Db 2383 ACTCCACATGCCCTCTCTTCTTTCGATCCCAACCGCCACAGGCGCTCTGTGCTTCTCT 2324

QY 111 CGAGGGGCGGTGTGATGTCTCCAGTATGTTCCGCCAGCGCTTTCCGACCCCTTCTGGA 170
Db 2233 CGAGGGGCGGTGTGATGTCTCCAGTATGTTCCGCCAGCGCTTTCCGACCCCTTCTGGA 2264

QY 171 CCAAGCGCCAAAGAGTGCAGCCAGAGAGAGGGGGCTCACTCTTATCTCTGGCGACCC 230
Db 2263 CCAAGCGCCAAAGAGTGCAGCCAGAGAGAGGGGGCTCACTCTTATCTCTGGCGACCC 2204

QY 231 ACTGCAAGAGCGCGCTCTCCAGACTTAAATGTATCACCATAACCTGTGAGGGG 290
Db 2203 ACTGCAAGAGCGCGCTCTCCAGACTTAAATGTATCACCATAACCTGTGAGGGG 2144

QY 291 ACCCAATCTGACTCTTCCCGCCTTGGGACATCGACGCGGGAAGCAGTGCAGCCCA 350
Db 2143 ACCCAATCTGACTCTTCCCGCCTTGGGACATCGACGCGGGAAGCAGTGCAGCCCA 2084

QY 351 GGCCTGGGCCAGAGAGCTCCAGGAAGGCGCACTGAGCGCTGTGGCGGAGGCTCGGAC 410
Db 2083 GGCCTGGGCCAGAGAGCTCCAGGAAGGCGCACTGAGCGCTGTGGCGGAGGCTCGGAC 2024

QY 411 ATCCGACGACACAGGGAAGTCTCCTGGGGCGATCTGTAAATAAACCTTTTCTTTT 470
Db 2023 ATCCGACGACACAGGGAAGTCTCCTGGGGCGATCTGTAAATAAACCTTTTCTTTT 1964

QY 471 GTTTTAAAAA 482
Db 1963 GTTTTAAAAA 1952

RESULT 5

US-10-187-657-3

; Sequence 3, Application US/10187657
; Publication No. US20030068311A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Azimzal, Yalda
; TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
; FILE REFERENCE: PV-0009 CIP
; CURRENT APPLICATION NUMBER: US/10/187, 657
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: PCT/US00/07817
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/139, 565
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 3
; LENGTH: 346
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030068311A1 1752794F6
; NAME/KEY: unsure
; LOCATION: 91, 188, 206, 291, 337
; OTHER INFORMATION: a, t, c, g, or other
US-10-187-657-3

Query Match 53.4%; Score 266; DB 14; Length 346;
Best Local Similarity 93.8%; Pred. No. 2.6e-70;
Matches 318; Conservative 0; Mismatches 15; Indels 6; Gaps 4;

QY 1 CTCTAGCGTGCCTCTGCGCGCTCCCGCTAGGCTCTCCGCGCTACACCACTCAGTACG 60
Db 13 CTCTAGCGTGCCTCTGCGCGCTCCCGCTAGGCTCTCCGCGCTACACCACTCAGTACG 71

QY 61 AGCGCCGCTCTGCGTCAAGCTGCAACCGCGCTCTGTGCTTCTCTCGAGGGGCC 120
Db 72 -GCGCGCGCTCTGCGTCAAGCTGCAACCGCGCTCTGTGCTTCTCTCGAGGGGCC 130

QY 121 GTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTCCGACCCCTTGGACCAAGGCC 180
Db 131 GTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTCCGACCCCTTGGACCAAGGCC 190

QY 181 AAGGACTGACCCAGAGAGAGAGGGGCTCACTCTTATCTCTGGCGA -CCCACTGACAA 239
Db 191 AAGGACTGACCCAGAGAGAGAGGGGCTCACTCTTATCTCTGGCGA -CCCACTGACAA 250

QY 240 GAGGCGCTCTCCAGACTTAAATGTATCACCATAACCTGTGAGGGGAGCCAACT 299
Db 251 GAGGCGCTCTCCAGACTTAAATGTATCACCATAACCTGTGAGGGGAGCCAACT 310

QY 300 GGAATCTTCCCGCTTGGGACATCGAGCGCGGGAAG 338
Db 311 GGAATCTT---CCCGCTTGGGAAATCGCANTCCGGGAAG 346

RESULT 6

US-10-187-657-9

; Sequence 9, Application US/10187657

```
/ Publication No. US20030068311A1
/ GENERAL INFORMATION:
/ APPLICANT: Lasek, Amy K.W.
/ APPLICANT: Baughn, Mariah R.
/ APPLICANT: Azimzai, Yalda
/ TITLE OF INVENTION: TRANSMEMBRANE PROTEIN DIFFERENTIALLY EXPRESSED IN CANCER
/ FILE REFERENCE: PV-0009 CIP
/ CURRENT APPLICATION NUMBER: US/10/187,657
/ PRIOR APPLICATION NUMBER: PCT/US00/07817
/ PRIOR FILING DATE: 2000-03-22
/ PRIOR APPLICATION NUMBER: 60/139,565
/ PRIOR FILING DATE: 1999-06-16
/ NUMBER OF SEQ ID NOS: 9
/ SOFTWARE: PERL Program
/ SEQ ID NO 9
/ LENGTH: 1594
/ TYPE: DNA
/ ORGANISM: Mus musculus
/ FEATURE:
/ NAME/KEY: misc_feature
/ OTHER INFORMATION: Incyte ID No. US20030068311A1 110769_Mm.1
US-10-187-657-9
```

```
Query Match      31.8%; Score 158.2; DB 14; Length 1594;
Best Local Similarity 65.7%; Pred. No. 2.2e-37;
Matches 328; Conservative 0; Mismatches 153; Indels 18; Gaps 6;
```

```
QY      2 TCTAGCGTGCCTCTGCGCGCTCCGCTAGGCTCTCCGCTCACCACCTAGTAGCA 61
Db      825 TCCAGCGTGCCTCTGCGCGCTCCGCTAGGCTCTCCGCTCACCACCTAGTAGCA 882
QY      62 GCGCCGCTTCTGCGTCAAGCGTCAAGCGCGCTGCTGCTGCTCTCTCGAGGGCGG 121
Db      883 GCGCCTCTTCTGCGTCAAGCGTCAAGCGCGCTGCTGCTGCTCTCTCGAGGGCGG 942
QY      122 TGTGAGTCTCCAGTATGTTGCGCGCGCTCTCTGCGACCTCTCTGAGCAAGCGCA 181
Db      943 TGTGATTTCTCACTATATCTCGCGCGCGCTCTCTCTCTCTGATCTAAGTGTCA 1002
QY      182 AGGACT--GCAGCCAGAGAGAGGGGGCTCACCCTCTTATCTCGCGGACCCACTGCACA 238
Db      1003 AAGACTGTAGCAACAGGCTAAAGGAACTCACCCTCTCACCCTCAACAACCCGCAACAG 1062
QY      239 AGCAGCGCGCTCTCCAGACTTAATATGATACCACTAACCTGTGAGGGGGACCCCAATC 298
Db      1063 AACAGTTGAAGAGTCCAGACTTAAA---TATTACCACTCTCTCTGAAAAATACTGACTC 1119
QY      299 TGGACTCTTCCCGCGCTTGGGACATGCAAGCGCGGGAAGCAGTCCCGCGAGCCCTG-G 357
Db      1120 CGGATTTCTAACCCCTCTTTGGGACCCCAATGACCTGAAGACTGTGTTAAGAGCGCGTGC 1179
QY      358 GCCAGGAGAGCTCCAGGAAGGCACTGAGCGCTGCGCGGCGGAGGCGCTCGACATCCGCA 417
Db      1180 AGGAGCATAGCTTAGTCTGGAAGACTGCTTGTCCCTGCGGAGCAGCAAGGGGGACCCCA 1239
QY      418 GGCACC-----AGGAAAGTCTCTCGGGCGCATCTGTAATAAACCCTTTTCTT 468
Db      1240 TGTGCCTTGTCTAGAGATGATTCTCAGGAGAACTGTAATAAACCCTTTTGTCTT 1299
QY      469 TTGTTTTTTAAAAAAAT 487
Db      1300 TTTTTTTTCTTCAAAAT 1318
```

```
RESULT 7
US-10-027-632-91970
/ Sequence 91970, Application US/10027632
/ Publication No. US20030204075A9
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ POLYMORPHISMS IN THE HUMAN GENOME
```

```
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ CURRENT FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 91970
/ LENGTH: 439
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(439)
/ OTHER INFORMATION: n = A,T,C or G
US-10-027-632-91970
```

```
Query Match      13.1%; Score 65.2; DB 15; Length 439;
Best Local Similarity 94.7%; Pred. No. 1.8e-09;
Matches 89; Conservative 0; Mismatches 3; Indels 2; Gaps 2;
```

```
QY      1 CTCTAGCGTGCCTCTGCGCGCTCCGCTAGGCTCTCCGCTCACCACCTAGTAGCA 60
Db      261 CTCTAGCGTGCCTCTGCGCGCTCCGCTAGGCTCTCCGCTCACCACCTAGTAGCA 319
QY      61 AGCGCGCGCTTCTGCGTCAAGCGTCAAGCGCGG 94
Db      320 -GGCGCGCTTCTGCGTCAAGCGTCAAGCGG 352
```

```
RESULT 8
US-10-123-155-10/C
/ Sequence 10, Application US/10123155
/ Publication No. US20030068794A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Beresini, Maureen
/ APPLICANT: DeForge, Laura
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Gao, Wei-Qiang
/ APPLICANT: Gerritsen, Mary E.
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Sherwood, Steven
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Watanabe, Colin K
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ ACIDS ENCODING THE SAME
/ FILE REFERENCE: P3330R1C30
/ CURRENT APPLICATION NUMBER: US/10/123,155
/ PRIOR APPLICATION REMOVED - See Palm or File Wrapper
/ NUMBER OF SEQ ID NOS: 550
/ SEQ ID NO 10
/ LENGTH: 594
```


TYPE: PRT
ORGANISM: Homo Sapien
US-10-123-155-10

Query Match 10.0%; Score 49.8; DB 14; Length 594;
Best Local Similarity 7.5%; Pred. No. 9.7e-05;
Matches 36; Conservative 183; Mismatches 260; Indels 0; Gaps 0;

```
QY 6 GCGTGCCGCTCTGCCGCTCCCGCTAGGCTCTCCGCGCTCAACCACTAGTACGAGCC 65
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 507 SYSYS.S.SWSSSYSSSDDY.CYCCYRHHCSDSYSYXY.CRCCTYT.SYRYDCHY 448
QY 66 CGCCTTCTGGGTACGCTGGCAACCGGCGCTCTGCTCTTCTCGAGAGGCGCTGCT 125
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 447 SCCSDYVCYSYSRYSYSYSWSYSYTDYCSYRCCCYSSSSSYSSSAY 388
QY 126 GAGTCTCCAGTATGTTCCGCCAGCGCTCTCGACCCCTTCTGACCAAGCGCCAAGA 185
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 387 STSSSSSSSSSYTSTNYC.T.CC...T.MCAABCTTTT...HSCC.SA.. 328
QY 186 CTGACGCCAGAGAGAGGGGCTCACTCTTATCTCGCGACCCACTGACACAGAGGC 245
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 327 A.M..YC.A.SYSYS.SSS.S.SYMR.HRA.SHYTRS..S.MCY.YM.Y.Y.Y.Y 268
QY 246 CGCTCTCCAGACTTAATATGATCAACCACTAGTGGAGGGGACCAATCTGACTC 305
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 SYCSRKT...TMTDM.T.T.MHY.KYB.HCHTKRAAT.MN.HTB...N.HBHB 208
QY 306 CTTCGCCCTTGGGACATCGACGCGGGAAGCAGTGCCCGGAGGCTGGGCCAGAG 365
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 207 .BH.H..B.H.HSNS...TTS....M.TTM.B.TBASH.RARC.STMM.T.MMM.H.S. 148
QY 366 AGCTCCAGAGAGGCACTGAGCGCTGCTGGCGAGGCGCTCGACATCCGACGACAG 425
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 147 Y..CCR..H.R..R.G...S.CT.HN.B.CYRY.RNGMY.HS....S.AA...ARH..D 88
QY 426 GGAAGTCTCTGGGCGGATCTGTAATAAACCTTTTCTTTGTTTAAAAAAA 484
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 87 H.DTYBH.YH.KNNAH.C.W.KKNB.S.TT.TBTSH.DBMGTYCD...T.HMM.M.RR 29
```

RESULT 9

US-10-146-731-10/c
Sequence 10, Application US/10146731
Publication No. US20030129692A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Gurney, Austin J.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C323
CURRENT APPLICATION NUMBER: US/10/146,731
CURRENT FILING DATE: 2002-05-15
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 10
LENGTH: 594
TYPE: PRT

ORGANISM: Homo Sapien
US-10-146-731-10

Query Match 10.0%; Score 49.8; DB 14; Length 594;
Best Local Similarity 7.5%; Pred. No. 9.7e-05;
Matches 36; Conservative 183; Mismatches 260; Indels 0; Gaps 0;

```
QY 6 GCGTGCCGCTCTGCCGCTCCCGCTAGGCTCTCCGCGCTCAACCACTAGTACGAGCC 65
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 507 SYSYS.S.SWSSSYSSSDDY.CYCCYRHHCSDSYSYXY.CRCCTYT.SYRYDCHY 448
QY 66 CGCCTTCTGGGTACGCTGGCAACCGGCGCTCTGCTCTTCTCGAGAGGCGCTGCT 125
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 447 SCCSDYVCYSYSRYSYSYSWSYSYTDYCSYRCCCYSSSSSYSSSAY 388
QY 126 GAGTCTCCAGTATGTTCCGCCAGCGCTCTTTCGACCCCTTCTGACCAAGCGCCAAGA 185
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 387 STSSSSSSSSSYTSTNYC.T.CC...T.MCAABCTTTT...HSCC.SA.. 328
QY 186 CTGACGCCAGAGAGAGGGGCTCACTCTTATCTCGCGACCCACTGACACAGAGGC 245
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 327 A.M..YC.A.SYSYS.SSS.S.SYMR.HRA.SHYTRS..S.MCY.YM.Y.Y.Y.Y 268
QY 246 CGCTCTCCAGACTTAATATGATCAACCACTAGTGGAGGGGACCAATCTGACTC 305
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 SYCSRKT...TMTDM.T.T.MHY.KYB.HCHTKRAAT.MN.HTB...N.HBHB 208
QY 306 CTTCGCCCTTGGGACATCGACGCGGGAAGCAGTGCCCGGAGGCTGGGCCAGAG 365
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 207 .BH.H..B.H.HSNS...TTS....M.TTM.B.TBASH.RARC.STMM.T.MMM.H.S. 148
QY 366 AGCTCCAGAGAGGCACTGAGCGCTGCTGGCGGAGGCGCTCGACATCCGACGACAG 425
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 147 Y..CCR..H.R..R.G...S.CT.HN.B.CYRY.RNGMY.HS....S.AA...ARH..D 88
QY 426 GGAAGTCTCTGGGCGGATCTGTAATAAACCTTTTCTTTGTTTAAAAAAA 484
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 87 H.DTYBH.YH.KNNAH.C.W.KKNB.S.TT.TBTSH.DBMGTYCD...T.HMM.M.RR 29
```

RESULT 10

US-10-140-472-10/c
Sequence 10, Application US/10140472
Publication No. US20030138888A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Gurney, Austin J.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C168
CURRENT APPLICATION NUMBER: US/10/140,472
CURRENT FILING DATE: 2002-05-06
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 10
LENGTH: 594
TYPE: PRT
ORGANISM: Homo Sapien

US-10-140-472-10

Query Match 10.0%; Score 49.8; DB 14; Length 594;
Best Local Similarity 7.5%; Pred. No. 9.7e-05;
Matches 36; Conservative 183; Mismatches 260; Indels 0; Gaps 0;

```
QY 6 GCGTGGCGCTCTGCGCGCTCCCGCTAGGCTCTCCGCGCTCACCCTCAGTACGAGCGC 65
Db 507 SYSYS.S.SWSYSYSSDDY.CYCCYRHCHSDSYSYY.CRCGYT.SYSRYDCHY 448
QY 66 CGCCTTCTGCTACGCTGCAACCGCGCTCTGCTCTCTCTCTCGAGGGCGCGGT 125
Db 447 SCCSDYCYCYSRYSYSYSYSYSYTDYCSYRCCCYSSSSSYSSSAY 388
QY 126 GAGTCTCCAGTATGTTCCGCCAGCGCTCTTGACACCTCTTGACCAAGCGCAAGA 185
Db 387 STSSSSSSSSYYTSTNYC.T.CC...T.MCAABCTTTT...HSCC.SA.. 328
QY 186 CTGACGCCAGAGAGAGGGGCTCACCCTTATCTCGGCGACCCACTGACAGAGGC 245
Db 327 A.M..YC.A.SYSYS.SS.S.SYMR.HRA.SHYTRS..S.MYCY.YM.Y.YY.YY 268
QY 246 CGCTCTCCAGACTTAAATGTATCACCACTAACCTGTGAGGGGAGCCCAATCTGACTC 305
Db 267 SYCSRKTM...TMTDM.T.T..MHMY.KYB.HCHTKCRAT.MN.HTB...N.HBHB 208
QY 306 CTTCGCCCTTGAGACATCGCAGCGCGGAGAGCAGTCCCGCGAGGCTGGCCAGAG 365
Db 207 .BH.H..B.H.HSNS...TTS...M.TTM.B.TBASH.RARC.STMM.T.MMM.H.S. 148
QY 366 AGCTCCAGAGAGGCACTGAGCGCTGCTGGCGGAGGCTCGACATCCGAGGACAG 425
Db 147 Y..CCR..H.R..R.G..S.CT.HN.B.CYRY.RNGMY.HS...S.AA...ARH..D 88
QY 426 GGAAGTCTCCGCGCGGATCTGTAATAAATTAACCTTTTCTTTTAAAAAA 484
Db 87 H.DTYBH.YH.KNNAH.C.W.KKNB.S.TT.TBTSH.DBMGYCDB...T.HMM.M.RR 29
```

RESULT 11

US-10-141-761-10/c

; Sequence 10, Application US/10141761

; Publication No. US20030148432A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C198
; CURRENT APPLICATION NUMBER: US/10/141,761
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo Sapien

US-10-141-761-10

Query Match 10.0%; Score 49.8; DB 14; Length 594;
Best Local Similarity 7.5%; Pred. No. 9.7e-05;
Matches 36; Conservative 183; Mismatches 260; Indels 0; Gaps 0;

```
QY 6 GCGTGGCGCTCTGCGCGCTCCCGCTAGGCTCTCCGCGCTCACCCTCAGTACGAGCGC 65
Db 507 SYSYS.S.SWSYSYSSDDY.CYCCYRHCHSDSYSYY.CRCGYT.SYSRYDCHY 448
QY 66 CGCCTTCTGCTACGCTGCAACCGCGCTCTGCTCTCTCTCTCGAGGGCGCGGT 125
Db 447 SCCSDYCYCYSRYSYSYSYSYSYTDYCSYRCCCYSSSSSYSSSAY 388
QY 126 GAGTCTCCAGTATGTTCCGCCAGCGCTCTTGACACCTCTTGACCAAGCGCAAGA 185
Db 387 STSSSSSSSSYYTSTNYC.T.CC...T.MCAABCTTTT...HSCC.SA.. 328
QY 186 CTGACGCCAGAGAGAGGGGCTCACCCTTATCTCGGCGACCCACTGACAGAGGC 245
Db 327 A.M..YC.A.SYSYS.SS.S.SYMR.HRA.SHYTRS..S.MYCY.YM.Y.YY.YY 268
QY 246 CGCTCTCCAGACTTAAATGTATCACCACTAACCTGTGAGGGGAGCCCAATCTGACTC 305
Db 267 SYCSRKTM...TMTDM.T.T..MHMY.KYB.HCHTKCRAT.MN.HTB...N.HBHB 208
QY 306 CTTCGCCCTTGAGACATCGCAGCGCGGAGAGCAGTCCCGCGAGGCTGGCCAGAG 365
Db 207 .BH.H..B.H.HSNS...TTS...M.TTM.B.TBASH.RARC.STMM.T.MMM.H.S. 148
QY 366 AGCTCCAGAGAGGCACTGAGCGCTGCTGGCGGAGGCTCGACATCCGAGGACAG 425
Db 147 Y..CCR..H.R..R.G..S.CT.HN.B.CYRY.RNGMY.HS...S.AA...ARH..D 88
QY 426 GGAAGTCTCCGCGCGGATCTGTAATAAATTAACCTTTTCTTTTAAAAAA 484
Db 87 H.DTYBH.YH.KNNAH.C.W.KKNB.S.TT.TBTSH.DBMGYCDB...T.HMM.M.RR 29
```

RESULT 12

US-10-142-885-10/c

; Sequence 10, Application US/10142885

; Publication No. US20030157604A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C248
; CURRENT APPLICATION NUMBER: US/10/142,885
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo Sapien

US-10-142-885-10

Query Match	10.0%;	Score 49.8;	DB 14;	Length 594;
Best Local Similarity	7.5%;	Pred. No. 9,7e-05;		
Matches 36;	Conservative 183;	Mismatches 260;	Indels 0;	Gaps 0

```

QY      6 GCGTGCCGCTCTGCCCCGCTCCCGCTAGAGCTCTCCGCGCTACCACTCAGTAGCAGCGC 65
Db      507 SYSYS.S.SMSSSYSSSDDY.CYCCYRHYHSDSISYSYY.CRCCTT.SYSRYDCHY 448
QY      66 CGCCTTCTGGGTACAGCTGGCAACCGGGCTCTGTGCTCTTCTCTGGAGGGCGCTGCT 125
Db      447 SCCCSDDYCYCSYSRYSISYSWSISYTDYCSYRCCCTIYSSSSYSSSAY 388
QY      126 GAGTCTCCAGTATGTTCCGCCACAGCGCTCTTCCGACCCTTTGACCAAAAGCGCCAAGA 185
Db      387 STSSSSSSSSYYTSTNYC.T.CC...T.MCAABCSTTTTTTTT.HSCC.SA. 328
QY      186 CTGACAGCAAGAGAGAGGGGCTCACCTCTTATCTCTGGCGACCCCACTGCACAGCAGGC 245
Db      327 A.M..YC.A.SYSYS.SSS.S.SYMR.HRA.SHYTRS.S.MYCY.YM.Y.YY.YY 268
QY      246 CGCTCTCCGACCTTAATAATGTATCACCACTAAGCTGTGAGGGGAGCCCAATCTGACTC 305
Db      267 SYCCSRKTM...TMTDM.T.T.MEMY.KYB.HCHTKCRAAT.MV.HTB...N.HBHB 208
QY      306 CTTCCCGCGCTTGACATCGCAGCGCGGAGAGCAGTGCCTCCGCGCTGCGGCAGAG 365
Db      207 .BH.H.B.H.HSNS...TTS...M.TTM.B.TBASH.RARC.STMM.T.MMM.H.S. 148
QY      366 AGCTCCAGGAAGGACGTAGCGCTGTGCGCGGAGGCTTCGACATCCGACGACCAAG 425
Db      147 Y..CCR..H.R..R.G...S.CT.HN.B.CYRY.RNGMY.HS...S.AA...ARH..D 88
QY      426 GGAAGCTCTCCTGGGGCATCTGTAAATAAACCTTTTTTCTTTGTTTAAAAAAA 484
Db      87 H.DTYBH.YH.KNNAHC.W.KNB.S.TT.TBTSH.DBMGTVCDB...T.HNM.M.RR 29

```

RESULT 13

US-10-158-790-10/c
; Sequence 10, Application US/10158790
; Publication No. US20030180879A1

GENERAL INFORMATION:

```

/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Beresini, Maureen
/ APPLICANT: DeForge, Laura
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Gao, Wei-Qiang
/ APPLICANT: Gerltsen, Mary E.
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Sherwood, Steven
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Watanabe, Colin K
/ APPLICANT: Wood, William
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ TITLE OF INVENTION: ACIDS ENCODING THE SAME
/ FILE REFERENCE: P3330R1C448
/ CURRENT APPLICATION NUMBER: US/10/156,790
/ CURRENT FILING DATE: 2002-05-30
/ Prior Application removed - See File Wrapper or Palm
/ NUMBER OF SEQ ID NOS: 550

```

```
; SEQ ID NO 10
;
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-158-790-10
```

Query Match 10.0%; Score 49.8; DB 14; Length 594;

Best Local Similarity 7.5%; Pred. No. 9.7e-05;
Matches 36; Conservative 183; Mismatches 260; Indels 0; Gaps 0;

```

QY      6 GCGTCCGCTCTGCCGCTCCCGCTAGGCTCTCTCCGCGTCAACCACTCAGTACGAGCC 65
Db      507 SYSYS.S.SWSSYSYSSDDY.CYCCYRHCSISYSYYY.CRCCTYT.SYSRYDCHY 448
QY      66 CGCCTTCGGGTCAAGCTGGCAACCGGCGCTCTGTGCCTCTTCTCGAGGGGCGCTGT 125
Db      447 SCCCDPYCYSYSRYYSYSYSYSWSYSSTYDYCSYRCCCYYSYSSSYSSSAY 388
QY      126 GAGTCTCCAGTATGTTCCGCCCAAGCGCTCTTCGACCCCTTTCGACCAACGCCCCAAGA 185
Db      387 STSSSSSSSSSYTSTNYC.T.CC....T.MCAABCTTTTTTTTTT..HSCC.SA.. 328
QY      186 CTGACCCAGAGAGAGGGGCTCACTTATCTCTCGGACCCCACTGCACAAAGCAGGC 245
Db      327 A.M..YC.A.SYSYSYS.SSS.S.SYMR.HRA.SHYTRS..S.MYCY.YM.Y.YY.YY 268
QY      246 CGCTCTCCAGACTTAATGATATCAACCACTAAGCTGTGAGGGGAGACCAATCTGACTC 305
Db      267 SYXCSRKT...TMTDM.T.T..MHY.KYB.HCHTKCRAAT.MN.HTB...N..HBHB 208
QY      306 CTTCGCCGCTTGGGACATGCAGAGCCGGAAAGCAGTGCCTCCGACGCTGGCCAGAG 365
Db      207 .BH.H..B.H.HSNS...TTS.....M.TTM.B.TBASH.PARC.STMM.T.MMM.H.S. 148
QY      366 AGCTCCAGGAAGGCACTGAGCGCTGTGCGGAGGCGCTGGACATCCGACGACACCAG 425
Db      147 Y..CCR..H.R..R.G..S.CT.HN.B.CYRY.RNGMY.HS...S.AA...ARH..D 88
QY      426 GGAAGTCTCTGGGCGCATCTGTAATAAACCTTTTCTTTTGTTTTAAAAAAA 484
Db      87 H.DTYBH.YH.KNNAH.C.W.KKNB.S.TT.TBISH.DBMGYCDB...T.HMM.M.RR 29

```

RESULT 14

US-10-137-871-10/c
; Sequence 10, Application US/10137871
; Publication No. US20030207350A1

GENERAL INFORMATION:

```

? APPLICANT: Baker, Kevin P.
? APPLICANT: Beresini, Maureen
? APPLICANT: DeForge, Laura
? APPLICANT: Desnoyers, Luc
? APPLICANT: Filvaroff, Ellen
? APPLICANT: Gao, Wei-Qiang
? APPLICANT: Geriltsen, Mary E.
? APPLICANT: Goddard, Audrey
? APPLICANT: Godowski, Paul J.
? APPLICANT: Gurney, Austin L.
? APPLICANT: Sherwood, Steven
? APPLICANT: Smith, Victoria
? APPLICANT: Stewart, Timothy A.
? APPLICANT: Tumas, Daniel
? APPLICANT: Watanabe, Colin K
? APPLICANT: Wood, William
? APPLICANT: Zhang, Zemin
? TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
? TITLE OF INVENTION: ACIDS ENCODING THE SAME
? FILE REFERENCE: P3330R1C153
? CURRENT APPLICATION NUMBER: US/10/137, 871
? CURRENT FILING DATE: 2002-05-03
? Prior Application removed - See Palm or File Wrapper
? NUMBER OF SEQ ID NOS: 550

```

SEQ ID NO 10

```

; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-137-871-10

```

Query Match	10.0%;	Score 49.8;	DB 15;	Length 594;
Best Local Similarity	7.5%;	Pred. No. 9.7e-05;		

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 19:35:23 ; Search time 1266.96 Seconds
(without alignments)
11737.841 Million cell updates/sec

Title: US-09-936-456-3
Perfect score: 498
Sequence: 1 ctcacgcgtgcgcgtctgcc.....aaaaaaataaagtcgacc 498

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query Match	Length	DB	ID	Description
C 1	457.8	91.9	486	9	AI660560	AI660560 we68b08.x
C 2	456	91.6	515	9	AI826629	AI826629 wk35e04.x
C 3	453	91.0	510	9	AI304327	AI304327 q057h06.x
C 4	446.6	89.7	537	9	AI821606	AI821606 nk08a11.x

C 5	435	87.3	446	9	AI984141	AI984141 wu21c02.x
C 6	431.4	86.6	466	9	AA573825	AA573825 nk08a11.s
C 7	423	84.9	434	9	AI991272	AI991272 wu41h04.x
C 8	414	83.1	414	9	AA593860	AA593860 nm19f03.s
C 9	409.4	82.2	405	9	AI983793	AI983793 wu20c09.x
C 10	405	81.3	405	9	AI274929	AI274929 q149c11.x
C 11	400.2	80.4	405	9	AI346155	AI346155 qp43f12.x
C 12	394.4	79.2	538	9	AI660493	AI660493 we67h02.x
C 13	387.4	77.8	573	14	CB854121	CB854121 ui-CF-DU1
C 14	381	76.5	381	9	AI281211	AI281211 qk58e08.x
C 15	378.2	75.9	737	14	CB305399	CB305399 ui-CF-EN1
C 16	335.8	67.4	690	12	BM977010	BM977010 ui-CF-EN1
C 17	335	67.3	348	9	AM050605	AM050605 w219b11.x
C 18	326.4	65.5	687	14	CB850544	CB850544 ui-CF-EN1
C 19	316.8	63.6	324	12	BM987789	BM987789 ui-H-CO0-
C 20	299.2	60.1	315	9	AA618335	AA618335 nq15g11.s
C 21	296	59.4	296	9	AI821178	AI821178 ne17a03.y
C 22	277	55.6	277	9	AI732165	AI732165 ne17a03.x
C 23	264.2	53.1	505	9	AI791844	AI791844 nk08a11.y
C 24	260	52.2	260	9	AM009962	AM009962 ws89h06.x
C 25	252	50.6	252	9	AI262416	AI262416 qk38e04.x
C 26	246.2	49.4	274	9	AA469031	AA469031 ne17a03.s
C 27	238.8	48.0	593	12	BM973444	BM973444 ui-CF-EC1
C 28	191.4	38.4	539	9	AI924216	AI924216 wu03b10.x
C 29	172.2	34.6	713	14	CB171450	CB171450 jtd602700
C 30	158.2	31.8	1594	11	AK018569	AK018569 Mus muscu
C 31	144.4	29.0	1309	11	AK008816	AK008816 Mus muscu
C 32	142	28.5	151	9	AI695625	AI695625 we50h04.x
C 33	138	27.7	408	12	BG938110	BG938110 labo11D08
C 34	135.4	27.2	602	12	BQ017315	BQ017315 ui-H-DT1-
C 35	94.4	19.0	378	10	BF917041	BF917041 il3-UT011
C 36	89	17.9	119	9	AI921289	AI921289 w022h06.x
C 37	79.4	15.9	538	13	BX517021	BX517021 BX517021
C 38	54.8	11.0	925	29	CNS0091P	AL053013 Drosophil
C 39	52.8	10.6	925	13	BX415111	BX415111 BX415111
C 40	51.4	10.3	925	29	CNS0091P	AL053013 Drosophil
C 41	46	9.2	479	14	H83064	H83064 yq46b05.r1
C 42	45.2	9.1	1201	13	BX357732	BX357732 BX357732
C 43	43.6	8.8	1000	13	BX407619	BX407619 BX407619
C 44	43.6	8.8	1201	13	BX360624	BX360624 BX360624
C 45	42.6	8.6	1269	12	BM807701	BM807701 AGENCOURT

ALIGNMENTS

RESULT 1
AI660560/c 486 bp mRNA linear EST 18-DEC-1999
LOCUS we68b08.x1 Soares Dieckgraeft_colon_NHCD Homo sapiens CDNA clone
DEFINITION IMAGE:2346231 3' similar to contains TARI.t3 TARI repetitive
element ;, mRNA sequence.

ACCESSION AI660560
VERSION AI660560.1 GI:4764130
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 486)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.

FEATURES
source 1..486
Insert Length: 143 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 447.
Location/Qualifiers

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2346231"
/tissue_type="colonic mucosa from 3 patients with Crohn's disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares_Dieckgraefe_colon_NHCD"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACCAATCTGAAGTGGAGGAGCGCGCGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieck@im.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bernaldo."

ORIGIN

Query Match 91.9%; Score 457.8; DB 9; Length 486;
Best Local Similarity 98.8%; Pred. No. 9e-80;
Matches 481; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

OY 5 AGCGTGGCGCTCTGCGCGCTCCCGCTAGGCTCTCCGCGTCAACCACTCACTAGTACGAGCG 64
DB 486 AGCGTGGCGCTCTGCGCGCTCCCGCTAGGCTCTCCGCGTCAACCACTCACTAGTACGAGCG 429
OY 65 CCGCCTTCTGCGTCAACCGCGCTCTGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 124
DB 428 CCGCCTTCTGCGTCAACCGCGCTCTGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 369
OY 125 TGAGTCTCCAGTATGTTGGCGCGCGCTCTGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 184
DB 368 TGAGTCTCCAGTATGTTGGCGCGCGCTCTGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 309
OY 185 ACTGAGCCAGGAGAGGGGCTCACTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 244
DB 308 ACTGAGCCAGGAGAGGGGCTCACTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 249
OY 245 CCGCTCTCCAGCTTAAATGATCACTCACTCACTCTGAGGGGAGCCCACTCTGACT 304
DB 248 CCGCTCTCCAGCTTAAATGATCACTCACTCACTCTGAGGGGAGCCCACTCTGACT 189
OY 305 CCTTCCCGCTTGGGACATCGAGCGCGGGAAGCAGTGCGCGCAGGCTGGGCGAGGA 364
DB 188 CCTTCCCGCTTGGGACATCGAGCGCGGGAAGCAGTGCGCGCAGGCTGGGCGAGGA 129
OY 365 GAGCTCCAGGAGGAGCAGTGAAGCGCTGCGCGAGGCGCTCGGACATCCGAGGACCA 424
DB 128 GAGCTCCAGGAGGAGCAGTGAAGCGCTGCGCGAGGCGCTCGGACATCCGAGGACCA 69
OY 425 GGGAAAGTCTCTGCGCGCATCTGTAATAAATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 484
DB 68 GGGAAAGTCTCTGCGCGCATCTGTAATAAATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 9
OY 485 AATAAAA 491
DB 8 AAAAAAA 2

RESULT 2
AI826629/c 515 bp mRNA linear EST 21-DEC-1999
LOCUS WK35e04.x1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:2417406 3',
DEFINITION mRNA sequence.
ACCESSION AI826629
VERSION AI826629.1 GI:5447300
KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 515)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 910 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 467.
Location/Qualifiers
1..515
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2417406"
/sex="male"
/tissue_type="normal prostate"
/lab_host="DH10B"
/clone_lib="NCI CGAP Pr22"
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from normal prostate bulk tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bernaldo."

FEATURES

source

ORIGIN
Query Match 91.6%; Score 456; DB 9; Length 515;
Best Local Similarity 99.6%; Pred. No. 2e-79;
Matches 478; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

OY 1 CTCTAGCGTGGCGCTCTGCGCGCTCCCGCTAGGCTCTCCGCGTCAACCACTCACTAGTACG 60
DB 478 CTCTAGCGTGGCGCTCTGCGCGCTCCCGCTAGGCTCTCCGCGTCAACCACTCACTAGTACG 420
OY 61 AGCGCGCTTCTGCGTCAACCGCGCTCTGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
DB 419 -GCGCGCTTCTGCGTCAACCGCGCTCTGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 361
OY 121 GTGTGAGTCTCCAGTATGTTGGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
DB 360 GTGTGAGTCTCCAGTATGTTGGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 301
OY 181 AAGGACTGAGCCAGGAGAGGGGCTCAGCTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240
DB 300 AAGGACTGAGCCAGGAGAGGGGCTCAGCTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 241
OY 241 CAGGCGCTCTCCAGACTTAAATGATCACTCACTCACTCTGAGGGGAGCCCAATCTG 300
DB 240 CAGGCGCTCTCCAGACTTAAATGATCACTCACTCACTCTGAGGGGAGCCCAATCTG 181
OY 301 GACTCTTCTCCCGCTTGGGACATCGAGGCGCGGGAAGCAGTGGCCCGCAGGCTGGGCC 360
DB 180 GACTCTTCTCCCGCTTGGGACATCGAGGCGCGGGAAGCAGTGGCCCGCAGGCTGGGCC 121
OY 361 AGGAGAGCTCCAGGAGGAGCAGTGAAGCGCTGTTGGGCGGAGGCGCTCGGACATCCGAGGC 420
DB 120 AGGAGAGCTCCAGGAGGAGCAGTGAAGCGCTGTTGGGCGGAGGCGCTCGGACATCCGAGGC 61


```

QY      421 ACCAGGGAAGTCTCTGGGGCGATCTGTAAATAAACCCTTTTCTTTGTTTTTAA 480
Db      60 ACCAGGGAAGTCTCTGGGGCGATCTGTAAATAAACCCTTTTCTTTGTTTTTAA 1

RESULT 3
AI304327/c      510 bp      mRNA      linear      EST 01-FEB-1999
LOCUS      q557h06.x1 NCI_CGAP Co8 Homo sapiens cDNA clone IMAGE:1912667 3,
DEFINITION      similar to contains TAR1.t3 TAR1 repetitive element ;, mRNA
sequence.
ACCESSION      AI304327
VERSION      AI304327.1 GI:3988016
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 510)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL      Tumor Gene Index
COMMENT      Unpublished (1997)
CONTACT      Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1487 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 476.
FEATURES
source
1. 510
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1912667"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI CGAP Co8"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo. "
ORIGIN

Query Match      91.0%; Score 453; DB 9; Length 510;
Best Local Similarity 99.6%; Pred. No. 7.7e-79;
Matches 475; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

```

```

Db      297 AAGACTGCAGCCAGAGAGAGGGGGCTCACTTATCTTCGGCAGCCACTGCACAG 238
QY      241 CAGCGCGCTCTCCAGACTTAAATGTATCACCACTAACCCTGTGAGGGGAGCCAACTGTG 300
Db      237 CAGCGCGCTCTCCAGACTTAAATGTATCACCACTAACCCTGTGAGGGGAGCCAACTGTG 178
QY      301 GACTCTTCCCGCCTTGGGACATCGCAGGCGGGAAGAGTGCCTCCGCAAGCCTGGGCC 360
Db      177 GACTCTTCCCGCCTTGGGACATCGCAGGCGGGAAGAGTGCCTCCGCAAGCCTGGGCC 118
QY      361 AGGAGAGCTCCAGAGAGGGGCACTGAGCGCTGCTGGCGGAGGCTCGGACATCCGACAGC 420
Db      117 AGGAGAGCTCCAGAGAGGGGCACTGAGCGCTGCTGGCGGAGGCTCGGACATCCGACAGC 58
QY      421 ACCAGGGAAGTCTCTGGGGCGATCTGTAAATAAACCCTTTTCTTTGTTTTT 477
Db      57 ACCAGGGAAGTCTCTGGGGCGATCTGTAAATAAACCCTTTTCTTTGTTTTT 1

RESULT 4
AI821606/c      537 bp      mRNA      linear      EST 13-DEC-1999
LOCUS      nk08a11.x5 NCI_CGAP Co2 Homo sapiens cDNA clone IMAGE:1012892 3,
DEFINITION      similar to contains TAR1.t3 TAR1 MER22 repetitive element ;, mRNA
sequence.
ACCESSION      AI821606
VERSION      AI821606.1 GI:5440685
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 537)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL      Tumor Gene Index
COMMENT      Unpublished (1997)
CONTACT      Other ESTs: nk08a11.y5
CONTACT: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc., David B. Krizman,
Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: National Cancer Institute, Cancer Genome
Anatomy Project (CGAP), Tumor Gene Index
This read has been verified (found to hit its original self in the
correct orientation)
Insert Length: 741 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 459.
FEATURES
source
1. 537
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1012892"
/tissue_type="tumor"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="NCI_CGAP Co2"
/note="Organ: colon; Vector: Bluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Bulk colon villous adenoma. 5' adaptor sequence:
5' GAATTCGGCAGAG 3' 3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTTCTTTT 3' Average insert size: 1.1 kb."

```

ORIGIN

Query Match 89.7%; Score 446.6; DB 9; Length 537;
Best Local Similarity 98.7%; Pred. No. 1.4e-77;
Matches 471; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 1 CTCTAGCGTGCCGCTCTGCCCCGCTCCGCTAGAGCTCTCCGCGCTCACCACTCAGTACG 60
Db 476 CTCTAGCGTGCCGCTCTGCCCCGCTCCGCTAGAGCTCTCCGCGCTCACCACTCAGTACG 418
QY 61 AGCGCCGCTCTTGGGTCACGCTGGCAACCGGCGCTCTGCTCTCTCTCGAGGGGCG 120
Db 417 -GCGCCGCTCTTGGGTCACGCTGGCAACCGGCGCTCTGCTCTCTCTCGAGGGGCG 359
QY 121 GTGTAGTCTCCAGTATGTTCGGCCCAAGCGCTCTTGGACCCCTTGGACCAAGCGCC 180
Db 358 GTGTAGTCTCCAGTATGTTCGGCCCAAGCGCTCTTGGACCCCTTGGACCAAGCGCC 299
QY 181 AAGGACTGCAGCCAGAGAGAGAGGGGCTCAGCTCTTATCTTGGGCGACCCACTGCACAG 240
Db 298 AAGGACTGCAGCCAGAGAGAGAGGGGCTCAGCTCTTATCTTGGGCGACCCACTGCACAG 239
QY 241 CAGCGCGCTCTCCAGACTTAATATGATACCACTAACCTGTAGAGGGGACCCCAATCTG 300
Db 238 CAGCGCGCTCTCCAGACTTAATATGATACCACTAACCTGTAGAGGGGACCCCAATCTG 179
QY 301 GACTCTTCCCGCTTGGGACATGCAGCGCCGGGAAGCAGTGCCTCCGCAAGCCTGAGCC 360
Db 178 GACTCTTCCCGCTTGGGACATGCAGCGCCGGGAAGCAGTGCCTCCGCAAGCCTGAGCC 119
QY 361 AGGAGAGCTCCAGAGAGGACACTGAGCGCTGTGGCGGAGGCTCGACATCCGACAGC 420
Db 118 AGGAGAGCTCCAGAGAGGACACTGAGCGCTGTGGCGGAGGCTCGACATCCGACAGC 59
QY 421 ACCAGGAAAGTCTCTGGGGCGATCTGTAATAACTTTTCTTTTCTTTT 477
Db 58 ACCAGGAAAGTCTCTGGGGCGATCTGTAATAACTTTTCTTTTCTTTTATGATTT 2

RESULT 5
AI984141/c

LOCUS AI984141 446 bp mRNA linear EST 27-OCT-1999
DEFINITION WU21C02.x1 Soares Dieckgraebe_colon_NHCD Homo sapiens cDNA clone
IMAGE:2520674 3', mRNA sequence.

ACCESSION AI984141.
VERSION AI984141.1 GI:5811360
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 446)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40up from Glibco.

FEATURES
source

1..446
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2520674"
/tissue_type="colonic mucosa from 3 patients with Crohn's
disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares Dieckgraebe_colon_NHCD"
/note="Organ: colon; Vector: pT7T3D-Fac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st

ORIGIN

Query Match 87.3%; Score 435; DB 9; Length 446;
Best Local Similarity 99.8%; Pred. No. 2.7e-75;
Matches 446; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTACCATCTGAGTGGAGCGGCCCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Tissue samples
provided by Dr. Brian Dieckgraebe (Washington University,
dieck@im.wustl.edu); colonic mucosa represents a range of
disease involvement from moderate to severe Crohn's
disease; samples include both perforating (fistulas) and
non-perforating samples. Library constructed by Bento
Soares and M. Fatima Bonaldo.

QY 34 GCTCCTCCGCGCTCACCACTCAGTACGAGCGCCGCTTCTGGTCAACGCTGGCAACCGGC 93
Db 446 GCTCCTCCGCGCTCACCACTCAGTACG-GCGCCGCTTCTGGTCAACGCTGGCAACCGGC 388
QY 94 GTCTGTGCTCTTCTCTGGAGGGGCGGTGTGATCTCCAGTATGTTCCGCCACGCT 153
Db 387 GTCTGTGCTCTTCTCTGGAGGGGCGGTGTGATCTCCAGTATGTTCCGCCACGCT 328
QY 154 CTTCGACCCCTTCTGACCAAGCGCCAAAGACTGACGACGAGAGAGGGGCTCACCT 213
Db 327 CTTCGACCCCTTCTGACCAAGCGCCAAAGACTGACGAGAGAGAGGGGCTCACCT 268
QY 214 CTTATCTCGGCGACCACTGACAGAGGCGCTCTCCCAAGCTTAATATGATCACC 273
Db 267 CTTATCTCGGCGACCACTGACAGAGGCGCTCTCCCAAGCTTAATATGATCACC 208
QY 274 ACTAACCTGTGAGGGGGAACCAATCTGACTCTTCCCGCTTGGAGACATCGAGGCG 333
Db 207 ACTAACCTGTGAGGGGGAACCAATCTGACTCTTCCCGCTTGGAGACATCGAGGCG 148
QY 334 GGAAGCAGTCCCGCCAGGCTGCGCCAGAGAGCTCCAGAAAGGCACTGAGGCTGCT 393
Db 147 GGAAGCAGTCCCGCCAGGCTGCGCCAGAGAGCTCCAGAAAGGCACTGAGGCTGCT 88
QY 394 GGGCGAGGCTTGGACATCCGACAGGCAACGAGAAAGTCTCTGGGGCGATCTGTAAT 453
Db 87 GGGCGAGGCTTGGACATCCGACAGGCAACGAGAAAGTCTCTGGGGCGATCTGTAAT 28
QY 454 AAACCTTTTCTTTTCTTTTAA 480
Db 27 AAACCTTTTCTTTTCTTTTAA 1

RESULT 6
AA573825/c

LOCUS AA573825 466 bp mRNA linear EST 12-SEP-1997
DEFINITION NK08a11.s1 NCI_CGAP_Co2 Homo sapiens cDNA clone IMAGE:1012892 3',
mRNA sequence.

ACCESSION AA573825
VERSION AA573825.1 GI:2348340
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS

1 (bases 1 to 466)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Stratagene, Inc., David B. Krizman,

Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution by: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrp/image/image.html
Insert Length: 741 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 456.

FEATURES

source

1. .466
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1012892"
/tissue_type="tumor"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="NCI CGAP Co2"
/note="Organ: colon; Vector: Bluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Bulk colon villous adenoma. 5' adaptor sequence:
5' GAATTCGGACGAG 3' 3' adaptor sequence: 5'
CTCAGTTTCTTTTCTTTTCTTTT 3' Average insert size: 1.1 Kb."

ORIGIN

Query Match

Best Local Similarity 99.1%; Pred. No. 1.4e-74;
Matches 465; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

QY 9 TGCCGCTTGCCCGCTCCCGCTAGAGTCTCCGCGCTCACCACTAGTACGAGCGCCG 68
Db 466 TGCCGCTTGCCCGCTCCCGCTAGAGTCTCCGCGCTCACCACTAGTACGAGCGCCG 409
QY 69 CTTCTGGGTCAAGCTGGCAACCGCGCTGTGCTCTTCTCGGAGGGCGCTGTGAG 128
Db 408 CTTCTGGGTCAAGCTGGCAACCGCGCTGTGCTCTTCTCGGAGGGCGCTGTGAG 349
QY 129 TCTCCAGTATGTTCCGCGCAGCGCTTTCGACACCTTCTGACCAAGCGCCAGGACTG 188
Db 348 TCTCCAGTATGTTCCGCGCAGCGCTTTCGACACCTTCTGACCAAGCGCCAGGACTG 289
QY 189 CAGCCAGAGAGAGGGGCTCACCTCTTATCTCGCGCAGCCCACTGCACAAGAGGCGCG 248
Db 288 CAGCCAGAGAGAGGGGCTCACCTCTTATCTCGCGCAGCCCACTGCACAAGAGGCGCG 229
QY 249 TCTCCAGACTTAAATGATATCACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTT 308
Db 228 TCTCCAGACTTAAATGATATCACCACTAACCTGTGAGGGGAGCCCAATCTGACTCTT 169
QY 309 CCCCCTTGGGACATCGCAGAGCGCGGAAAGCAGTGCCTCCGCGCAGGCTGGCCAGAGAGC 368
Db 168 CCCCCTTGGGACATCGCAGAGCGCGGAAAGCAGTGCCTCCGCGCAGGCTGGCCAGAGAGC 109
QY 369 TCCAGGAAGGGCACTGAGCGCTGTGCGCGAGGCGCTCGACATCCGAGGCAAGGGA 428
Db 108 TCCAGGAAGGGCACTGAGCGCTGTGCGCGAGGCGCTCGACATCCGAGGCAAGGGA 50
QY 429 AAGTCTCTGGGGCGATGTGTAATAAACCTTTTCTTTTCTTTT 477
Db 49 AAGTCTCTGGGGCGATGTGTAATAAACCTTTTCTTTTCTTTT 1

RESULT 7

AI991272/c

LOCUS AI991272 434 bp mRNA linear EST 09-MAR-2000
DEFINITION wu41h04.x1 Soares_Dieckgraefe_colon_NHCD Homo sapiens cDNA clone
IMAGE:2522647 3', mRNA sequence.

ACCESSION AI991272

VERSION AI991272.1 GI:5838177
KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
1 (bases 1 to 434)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaaps-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 558 Std Error: 0.00
Seq primer: -40UP from Gibco.

FEATURES

source

1. .434
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2522647"
/tissue_type="colonic mucosa from 3 patients with Crohn's
disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares_Dieckgraefe_colon_NHCD"
/note="Organ: colon; Vector: pRT3D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dt) primer (5'
TGTTACCAATCTGAAGTGGAGCGGCGCCCTTTTCTTTTCTTTT 3'),
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pRT3D vector. Library
went through one round of normalization. Tissue samples
provided by Dr. Brian Dieckgraefe (Washington University,
dieck@im.wustl.edu); colonic mucosa represents a range of
disease involvement from moderate to severe Crohn's
disease; samples include both perforating (fistulas) and
non-perforating samples. Library constructed by Bento
Soares and M. Fatima Bonaldo."

ORIGIN

Query Match

Best Local Similarity 99.8%; Pred. No. 6.2e-73;
Matches 434; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 44 GCTTACCACCTAGTACGAGGCGCGCTTGTGGTACCGCTGGCAACCGCGCTGTGCCC 103
Db 434 GCTTACCACCTAGTACGAGGCGCGCTTGTGGTACCGCTGGCAACCGCGCTGTGCCC 376
QY 104 TCTTCTCGGAGGGCGCGTGTGAGTCTCAGTATGTTTGGCCCGGCTCTTGGCACCC 163
Db 375 TCTTCTCGGAGGGCGCGTGTGAGTCTCAGTATGTTTGGCCCGGCTCTTGGCACCC 316
QY 164 TTCTGGACCAAGCGCCCAAGACTGCAGCCAGAGAGAGGGGCTCACTTATCTCG 223
Db 315 TTCTGGACCAAGCGCCCAAGACTGCAGCCAGAGAGAGGGGCTCACTTATCTCG 256
QY 224 GCGACCCACTGCACAAGAGCGCGCTCTCCAGACTTAAATGTATCACCATACTGT 283
Db 255 GCGACCCACTGCACAAGAGCGCGCTCTCCAGACTTAAATGTATCACCATACTGT 196
QY 284 GAGGGGAGCCCAATCTGACTCTTCCCGCGCTTGGACATCGCAGGCGGGAAGCACTG 343
Db 195 GAGGGGAGCCCAATCTGACTCTTCCCGCGCTTGGACATCGCAGGCGGGAAGCACTG 136
QY 344 CCGCCAGGCGCTGGCCAGAGAGCTCCAGGAAGGCACTGAGCGCTGTGGCGGAGGC 403
Db 135 CCGCCAGGCGCTGGCCAGAGAGCTCCAGGAAGGCACTGAGCGCTGTGGCGGAGGC 76
QY 404 CTGGACATCGCAGGACCAAGGAAGTCTCTGGGGGATCTGTAATAAACCTTTT 463
Db 75 CTGGACATCGCAGGACCAAGGAAGTCTCTGGGGGATCTGTAATAAACCTTTT 16
QY 464 TTCTTTTGTTTT 478
Db 15 TTCTTTTGTTTT 1

RESULT 8
AA593860/c

LOCUS 414 bp mRNA linear EST 25-SEP-1997
DEFINITION un19f03.s1 NCI_CGAP_Col12 Homo sapiens cDNA clone IMAGE:1084349 3',
mRNA sequence.

ACCESSION AA593860
VERSION AA593860
KEYWORDS GI:2408538
SOURCE EST.

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 414)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: Stratagene, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 1204 Std Error: 0.00
Seq primer: -40m3 fwd. RT from Amersham
High quality sequence stop: 406.
Location/Qualifiers

FEATURES
source

1..414
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1084349"
/sex="mixed"
/tissue_type="colon tumor"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="NCI CGAP Col12"
/note="Organ: colon; Vector: Bluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pooled colon tumors. 5' adaptor sequence: 5'
GAATTCGGACGAG 3' 3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."

ORIGIN

Query Match 83.1%; Score 414; DB 9; Length 414;
Best Local Similarity 100.0%; Pred. No. 3.7e-71;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 63 CGCCGCTTCTGGGTCAACGCTGGCAACCGCGCTCTGCTCTTCTCGAGGGGCGGT 122
DB 414 CGCCGCTTCTGGGTCAACGCTGGCAACCGCGCTCTGCTCTTCTCGAGGGGCGGT 355
QY 123 GGTAGTCTCCAGTATGTTGGGCCCAAGCGCTCTTTCGACCCCTTTCGACCAAGCGCAA 182
DB 354 GGTAGTCTCCAGTATGTTGGGCCCAAGCGCTCTTTCGACCCCTTTCGACCAAGCGCAA 295
QY 183 GGACTGACCCAGAGAGAGAGGGGGCTCACTTTATCTCTGGGCAACCACTGACAAAGCA 242
DB 294 GGACTGACCCAGAGAGAGAGGGGGCTCACTTTATCTCTGGGCAACCACTGACAAAGCA 235
QY 243 GGCCGCTCTCCAGACTTAATGTATCACCACTAACCCTGTGAGGGGGAGCCCAATCTGGA 302
DB 234 GGCCGCTCTCCAGACTTAATGTATCACCACTAACCCTGTGAGGGGGAGCCCAATCTGGA 175
QY 303 CTCCTTCCCGCCTTGGGACATCGAGGGCCGGGAAGCAGTGCCCGCCAGGCTGGGCCAG 362
DB 174 CTCCTTCCCGCCTTGGGACATCGAGGGCCGGGAAGCAGTGCCCGCCAGGCTGGGCCAG 115

QY 363 GAGAGCTCCAGGAAGGCACTGAGCGCTGCTGGCCGAGGCGCTTCGACATCCGACAGCAC 422
DB 114 GAGAGCTCCAGGAAGGCACTGAGCGCTGCTGGCCGAGGCGCTTCGACATCCGACAGCAC 55
QY 423 CAGGAAAGTCTCCTGGGGCGATCTGTAATAAACCCTTTTCTTTGTTTTT 476
DB 54 CAGGAAAGTCTCCTGGGGCGATCTGTAATAAACCCTTTTCTTTGTTTTT 1

RESULT 9
AI983793/c

LOCUS 411 bp mRNA linear EST 27-OCT-1999
DEFINITION wu20c09.x1 Soares Dieckgraefe_colon_NHCD Homo sapiens cDNA clone
IMAGE:2520592 3', mRNA sequence.

ACCESSION AI983793
VERSION AI983793
KEYWORDS GI:5811012
SOURCE EST.

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 411)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 400.
Location/Qualifiers

FEATURES
source

1..411
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2520592"
/tissue_type="colonic mucosa from 3 patients with Crohn's
disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares Dieckgraefe_colon_NHCD"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dt) primer [5'
TGTTACCAATCTGAAGTGGAGAGCGCGCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Tissue samples
provided by Dr. Brian Dieckgraefe (Washington University,
dieck@im.wustl.edu); colonic mucosa represents a range of
disease involvement from moderate to severe Crohn's
disease; samples include both perforating (fistulas) and
non-perforating samples. Library constructed by Bento
Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 82.2%; Score 409.4; DB 9; Length 411;
Best Local Similarity 99.8%; Pred. No. 2.9e-70;
Matches 410; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 72 CTGGGTACGCTGGCAACCGGCGTCTGTGCTCTTCTCGAGGGCGGTGTGAGTCT 131
DB 411 CTGGGTACGCTGGCAACCGGCGTCTGTGCTCTTCTCGAGGGCGGTGTGAGTCT 352
QY 132 CCAGTATGTTGGCCAGCGCTCTTGCAACCTTTCGACCAAGCGCCAGAGACTGCAG 191
DB 351 CCAGTATGTTGGCCAGCGCTCTTGCAACCTTTCGACCAAGCGCCAGAGACTGCAG 292
QY 192 CCAGAGAGAGGGGCTCACCCTTATCTCTGGCGAACCACTGCACAAGCAGCGGCTCT 251
DB 291 CCAGAGAGAGGGGCTCACCCTTATCTCTGGCGAACCACTGCACAAGCAGCGGCTCT 232

QY	252	CCCAGACTTAAATATGATATCACCACCTAACCTGTGAGGGGGACCCCAATCTGGACTCCTTCCC	311
Db	231	CCCAGACTTAAATATGATATCACCACCTAACCTGTGAGGGGGACCCCAATCTGGACTCCTTCCC	172
QY	312	CGCCTTGGGACATCCGAGGCCCGGGAAGCAGTGCCTCCGCCAGGCTTGGGCCAGAGAGCTCC	371
Db	171	CGCCTTGGGACATCCGAGGCCCGGGAAGCAGTGCCTCCGCCAGGCTTGGGCCAGAGAGCTCC	112
QY	372	AGGAAGGGCACTGAGCGCTGTGCGCGCAGAGCCTCGGACATCCGACAGCACCCAGGGAAG	431
Db	111	AGGAAGGGCACTGAGCGCTGTGCGCGCAGAGCCTCGGACATCCGACAGCACCCAGGGAAG	52
QY	432	TCTCCTGGGGCGATCTGTAAATAAACCTTTTCTTTCTTTTAAAAA	482
Db	51	TCTCCTGGGGCGATCTGTAAATAAACCTTTTCTTTCTTTTAAAAA	1
RESULT 10			
AI274929/c			
LOCUS	AI274929	405 bp	mRNA linear EST 29-JAN-1999
DEFINITION	q149c11.x1 NCI CGAP Co8 Homo sapiens cDNA clone IMAGE:1875668 3'		

ACCESSION	AI274929	
VERSION	AI274929.1	GI:3897203
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	

REFERENCE
AUTHORS
TITLE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
1 (bases 1 to 405)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute Cancer Genome Anatomy Project (CGAP).

**JOURNAL
COMMENT**

**Tumor Gene Index
Unpublished (1997)**

Contact: Robert Strausberg, Ph.D.
Email: csapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1458 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop:395.

FEATURES	Location/Qualifiers
source	1. .405

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1875668"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI CGAP_C08"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo. "

```

ORIGIN

Query Match	81.3%;	Score 405;	DB 9;	Length 405;
Best Local Similarity	100.0%;	Pred. No. 2.1e-69;		
Matches 405; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
73	TGGGTACCGCTGGCAACCGCGCTCTGTGCTCTTCTCTCGAGAGGCGCGTGTGATCTC	132		

Db	405	TGGGTCA	CGCTG	CAACCG	GGCGT	CTGTG	CTCTT	CTCCG	AGGGG	CCGTG	TGAGT	CTC	346			
QY	133	CAGTATG	TTCGG	CCAGC	CGCTCTT	TGCA	CCCTT	CTG	ACCAAA	CGCC	CAAG	ACTG	CAGC	192		
Db	345	CAGTATG	TTCGG	CCAGC	CGCTCTT	TGCA	CCCTT	CTG	ACCAAA	CGCC	CAAG	ACTG	CAGC	286		
QY	193	CAGGAG	AGAGG	GGGCT	CACTCTT	ATCT	CTGG	CGAC	CCCACT	GTG	CA	CAAG	CAGCG	CTCTC	252	
Db	285	CAGGAG	AGAGG	GGGCT	CACTCTT	ATCT	CTGG	CGAC	CCCACT	GTG	CA	CAAG	CAGCG	CTCTC	226	
QY	253	CCAG	ACTTAA	ATGTA	TACCA	CTAA	CCCTG	TAG	GGGGG	AGCC	CAATCT	GG	ACTCT	CTTCCC	312	
Db	225	CCAG	ACTTAA	ATGTA	TACCA	CTAA	CCCTG	TAG	GGGGG	AGCC	CAATCT	GG	ACTCT	CTTCCC	166	
QY	313	GCCTT	GGG	ACATCG	CAGG	CCCGG	GAAG	ACAGT	GC	CCCGC	CAGG	CTGGG	CCAG	AGAG	CTCCA	372
Db	165	GCCTT	GGG	ACATCG	CAGG	CCCGG	GAAG	ACAGT	GC	CCCGC	CAGG	CTGGG	CCAG	AGAG	CTCCA	106
QY	373	GGAAG	GGCA	CTG	AGCG	CTG	TGG	CGCG	AGG	CGCTCG	ACAT	CCG	CAG	CA	GGAAAGT	432
Db	105	GGAAG	GGCA	CTG	AGCG	CTG	TGG	CGCG	AGG	CGCTCG	ACAT	CCG	CAG	CA	GGAAAGT	46
QY	433	CTCCT	GGG	CGATCT	GTA	ATA	TAA	CC	TTTTTT	CTTTT	GT	TTTTT	477			
Db	45	CTCCT	GGG	CGATCT	GTA	ATA	TAA	CC	TTTTTT	CTTTT	GT	TTTTT	1			

RESULT	11
AI346155/c	
LOCUS	
DEFINITION	AI346155 405 bp mRNA linear EST 02-FEB-1999 gp43f12.x1 NCI CGAP C08 Homo sapiens cDNA clone IMAGE:1925807 3'
ACCESSION	sequence.
VERSION	AI346155
KEYWORDS	AI346155.1 GI:4083361
SOURCE	EST.
ORGANISM	Homo sapiens (human) Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 405)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1488 Std Error: 0.00.
Seq primer: -40UP from Gibco
High quality sequence stop: 375.

```
FEATURES      Location/Qualifiers
source        1. .405
```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1925807"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Co8"
/note="Organ: colon; Vector: pTTT3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned

```


into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match	80.4%;	Score 400.2;	DB 9;	Length 405;
Best Local Similarity	99.3%;	Pred. No. 1.9e-68;		
Matches 402; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

QY	73	TGGGTACCGCTGGCAACCGGGCGTCTGTGCTCTTCTCTGGAGGGGCCGTGTGAGTCTC	132
Db	405	TGGGTACCGCTGGCAACCGGGCGTCTGTGCTCTTCTCTGGAGGGGCCGTGTGAGTCTC	346
QY	133	CAGTATGTTCCGGCCCAAGGCTCTTCGCACCTTCTTGACCAAAAGCCCAAGACTGCAGC	192
Db	345	CAGTATGTTCCGGCCCAAGGCTCTTCGCACCTTCTTGACCAAAAGCCCAAGACTGCAGC	286
QY	193	CAGAGAGAGGGGGCTCACTCTTATTCCTCGGCGACCCACTGCACAAGCAGCGGCTCTC	252
Db	285	CAGAGAGAGGGGGCTCACTCTTATTCCTCGGCGACCCACTGCACAAGCAGCGGCTCTC	226
QY	253	CCAGACTTAAATGTATCAACCACTAACCTGTGAGGGGGACCCCACTCTGACTCTTCCCC	312
Db	225	CCAGACTTAAATGTATCAACCACTAACCTGTGAGGGGGACCCCACTCTGACTCTTCCCC	166
QY	313	GCCTTGGGACATCGCAGGCCGGGAAGCAGTGC CGGCCAGGCTTGGGCCAGAGAGCTCCA	372
Db	165	GCCTTGGGACATCGCAGGCCGGGAAGCAGTGC CGGCCAGGCTTGGGCCAGAGAGCTCCA	106
QY	373	GGAAGGGCACTGAGCGCTGTGCGCGAAGGCTCTGGACATCCGCAGGACCAAGGAAAGT	432
Db	105	GGAAGGGCACTGAGCGCTGTGCGCGAAGGCTCTGGACATCCGCAGGACCAAGGAAAGT	46
QY	433	CTCCTGGGCGCATCTGTAAATAAACCTTTTTTCTTTTGTTTTTT 477	
Db	45	CTCCTGGGCGCATCAAAAATAAACCTTTTTTCTTTTGTTTTTT 1	

RESULT 12	LOCUS	DEFINITION
AI660493/c	AI660493	538 bp mRNA linear EST 18-DEC-1999
	we67h02.x1 Soares_Dieckgraefe_colon_NHCD Homo sapiens cdna clone	
	IMAGE:2346195.3 similar to contains TAR1.t3 TAR1 repetitive	
	element ;, mRNA sequence.	

FEATURES

SOURCE

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2346195"
/tissue_type="colonic mucosa from 3 patients with Crohn's
disease"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares Dieckgraefe colon NHCD"

```

/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo (dT) primer [5' GTTACCAATCTGAGTGGAGGCGCCGCTCTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgreffe (Washington University, dieck@em.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaído "

ORIGIN

Query Match	79.28;	Score 394.4;	DB 9;	Length 538;
Best Local Similarity	95.18;	Pred. No. 2.3e-67;		
Matches 407;	Conservative	0;	Mismatches 21;	Indels 0;
				Gaps 0;

QY	51	ACTCAGTACGAGGCGCGCCTTCTGGGTCA	C	CGGTGGCAACCGGCGTCTGTGCTCTTCT	110
Db	428	ACTCCACATGCCCTCTCTTCTTTCGATCC	CA	CGCCACAGGCGTCTGTGCTCTTCTCT	369
QY	111	CGAGGGGCGCGTGTGAGTCTCCAGTATG	T	TGCGCCAGGCGCTCTTGCACCCCTTCTGA	170
Db	368	CGAGGGGCGCGTGTGAGTCTCCAGTATG	T	TGCGCCAGGCGCTCTTGCACCCCTTCTGA	309
QY	171	CCAAAGCGCCCAAGACTGCAGCCAGAGAG	AGGGGCTCACCTCTTATCTCGGCGACCC	230	
Db	308	CCAAAGCGCCCAAGACTGCAGCCAGAGAG	AGGGGCTCACCTCTTATCTCGGCGACCC	249	
QY	231	ACTGCACAAGCAGGCGCGCTCTCCAGACT	T	TAAATGTATACCACTAACCTGTGAGGGG	290
Db	248	ACTGCACAAGCAGGCGCGCTCTCCAGACT	T	TAAATGTATACCACTAACCTGTGAGGGG	189
QY	291	ACCCAATCTGGACTCCTTCCCGCGCTTG	GA	CATCGCAGGCGGGAAGCAGTGCCCGCCA	350
Db	188	ACCCAATCTGGACTCCTTCCCGCGCTTG	GA	CATCGCAGGCGGGAAGCAGTGCCCGCCA	129
QY	351	GGCCTGGGCGCAGAGAGCTCCAGAAAGG	CACTGAGCGCTGTGCGCAGGCGCTCGAC	410	
Db	128	GGCCTGGGCGCAGAGAGCTCCAGAAAGG	CACTGAGCGCTGTGCGCAGGCGCTCGAC	69	
QY	411	ATCCGAGGCACCAAGGAAAGTCTCTG	GGCGATCTGTAAATAAACCTTTTTCCTTT	470	
Db	68	ATCCGAGGCACCAAGGAAAGTCTCTG	GGCGATCTGTAAATAAACCTTTTTCCTTT	9	
QY	471	GTITTTTA	478		
Db	8	GTITTTTA	1		

RESULT 13					
CB854121/c					
LOCUS					
DEFINITION	CB854121	573 bp	mRNA	linear	EST 22-APR-2003
	UI-CF-DU1-aal-1-16-0-UI.a1		UI-CF-DU1 Homo	sapient	CDNA clone
	UI-CF-DU1-aal-1-16-0-UI 3'		mRNA sequence.		

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

1 (bases 1 to 573)
Bonald, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (3), 791-806 (1996)
97044477
8889548
Contact: McCray, PB

McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this CDNA
sequence: 533-573, >ALU (matched complement)
Seq primer: M13 FORWARD
POLYA=NO.

FEATURES
source

Location/Qualifiers
1. 573
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-DUI-aal-1-16-0-UI"
/tissue_type="Primary Lung Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-DUI"
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-CF-DUI is a normalized cDNA library containing the
following tissue(s): Primary Lung Epithelial Cells The
library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT7T3-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is Ggctgtagcgc.
TAG_SEQ=None found"

ORIGIN

Query Match 77.8%; Score 387.4; DB 14; Length 573;
Best Local Similarity 94.9%; Pred. No. 5.3e-66;
Matches 411; Conservative 0; Mismatches 21; Indels 1; Gaps 1;

QY 51 ACTCAGTACGAGCGCGCGCTTCTGGGTACGCTGGCAACCGCGCTGTGCTCTT-CC 109
DB 540 ACTCCACATGCCCCCTTCTTTCGATCCACCGCCACAGGCGTGTGCTCTTCCC 481
QY 110 TCGAGAGGCGCGTGTGATCTCCAGTATGTTGCGGCCAGCGCTCTTGCACCCCTTCTGG 169
DB 480 TCGAGAGGCGCGTGTGATCTCCAGTATGTTGCGGCCAGCGCTCTTGCACCCCTTCTGG 421
QY 170 ACCAAGCGCCCAAGGACTGACGCCAGAGAGAGAGGCGCTCACCCTTATCTCGCGCACC 229
DB 420 ACCAAGCGCCCAAGGACTGACGCCAGAGAGAGAGGCGCTCACCCTTATCTCGCGCACC 361
QY 230 CACTGCACAGCAGCGCGCTCTCCCAAGCTTAAATGTATCACCCTAACCCTGTGAGGGG 289
DB 360 CACTGCACAGCAGCGCGCTCTCCCAAGCTTAAATGTATCACCCTAACCCTGTGAGGGG 301
QY 290 GACCCCAATCTGACTCTCTCCCGCGCTTGGGACATCGCAGCGCGCGGAGCAGTCCCGCC 349
DB 300 GACCCCAATCTGACTCTCTCCCGCGCTTGGGACATCGCAGCGCGCGGAGCAGTCCCGCC 241
QY 350 AGGCTGGGCGCAGAGAGCTCCAGAAAGGCGCAGTACGCGCTGTGGCGCGGAGGCTCGGA 409
DB 240 AGGCTGGGCGCAGAGAGCTCCAGAAAGGCGCAGTACGCGCTGTGGCGCGGAGGCTCGGA 181
QY 410 CATCCGCGAGCAGCAGGAGAAAGTCTCCTGGGCGCATCTGTAAATAACCTTTTCTTT 469

DB 180 CATCCGCGAGGACACAGGAGAAAGTCTCCTGGGCGCATCTGTAAATAACCTTTTCTTT 121
QY 470 TGTTTTAAAAA 482
DB 120 TGTTTTAAAAA 108

RESULT 14
AI281211/c
LOCUS
DEFINITION
AI281211 381 bp mRNA linear EST 28-JAN-1999
qk58e08.x1 NCI CGAP Co8 Homo sapiens CDNA clone IMAGE:1873190 3',
similar to contains_TARI.t3 TARI repetitive element ;, mRNA
sequence.

ACCESSION
AI281211
VERSION
AI281211.1 GI:3919444
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 1422 Std Error: 0.00
Seq primer: -40up from Gibco
High quality sequence stop: 340.

FEATURES
source

Location/Qualifiers
1. 381
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1873190"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI CGAP Co8"
/note="Organ: Colon; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 76.5%; Score 381; DB 9; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.1e-64;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 CTGTGCTCTTCTCTCGAGGGGCGGTGTGAGTCTCCAGTATGTTCCGCCACGCTCTT 156
DB 381 CTGTGCTCTTCTCTCGAGGGGCGGTGTGAGTCTCCAGTATGTTCCGCCACGCTCTT 322
QY 157 CGACCCCTTGTGACCAAGCGCCCAAGACTGACGCCAGAGAGAGAGGGGCTCACCCTT 216
DB 321 CGACCCCTTGTGACCAAGCGCCCAAGACTGACGCCAGAGAGAGAGGGGCTCACCCTT 262
QY 217 ATCCTGGGCGACCCACTGACCAAGCAGCGCTCTCCAGACTTAAATGTATGACCACT 276
DB 261 ATCCTGGGCGACCCACTGACCAAGCAGCGCTCTCCAGACTTAAATGTATGACCACT 202

QY 277 AACCTGTGAGGGGACCCCAATCTGACTCTTCCCGCTTGGGACATCGAGCCGGGA 336
|||||
Db 201 AACCTGTGAGGGGACCCCAATCTGACTCTTCCCGCTTGGGACATCGAGCCGGGA 142
QY 337 AGCAGTCCCGCCGAGGCTTGGGACAGAGAGCTCCAGAGGCACTGAGCGCTGTGGC 396
|||||
Db 141 AGCAGTCCCGCCGAGGCTTGGGACAGAGAGCTCCAGAGGCACTGAGCGCTGTGGC 82
QY 397 GCGAGGCTCGGACATCCGACAGGCAAGAAAGTCTCTGGGGCGATCTGTAATAAA 456
|||||
Db 81 GCGAGGCTCGGACATCCGACAGGCAAGAAAGTCTCTGGGGCGATCTGTAATAAA 22
QY 457 CCTTTTCTTTCTTTTCTTTT 477
|||||
Db 21 CCTTTTCTTTCTTTTCTTTT 1

RESULT 15
CB305399
LOCUS 737 bp mRNA linear EST 04-MAR-2003
DEFINITION UI-CF-EN1-aed-m-17-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
UI-CF-EN1-aed-m-17-0-UI 3', mRNA sequence.
ACCESSION CB305399
VERSION CB305399.1 GI:28845910
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 737)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
The following repetitive elements were found in this cDNA
sequence: 29-138, >MR#SINE/MIR (matched complement) 195-290, >ALU
223-320, >ALU
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source
Location/Qualifiers
1..737
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-EN1-aed-m-17-0-UI"
/tissue_type="Primary Lung Cystic Fibrosis Epithelial
Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_id="UI-CF-EN1"
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site_1: Ecor I; Site_2: Not I;
UI-CF-EN1 is a normalized cDNA library containing the
following tissue(s): Primary Lung Cystic Fibrosis
Epithelial Cells. The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. First strand cDNA synthesis was primed with an

oligo-dT primer containing a Not I site. Double stranded
cDNA was ligated to an Ecor I adaptor, digested with Not
I, and cloned directionally into pT7T3-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CTGCTCAGGT.
TAG_TISSUE=Human Lung Epithelial Cell lines untreated LPS
6hr to LPS 24h
TAG_LIB=UI-CF-EN1
TAG_SEQ=CTGCTCAGGT"

ORIGIN
Query Match 75.9%; Score 378.2; DB 14; Length 737;
Best Local Similarity 92.9%; Pred. No. 3.1e-64;
Matches 395; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
QY 51 ACTCAGTACGAGCGCCGCTTCTGTGCTACGCTGCAACCGCGCTGTGCTCTTCT 110
|||||
Db 313 ACTCCACATGCGCTCTCTTCTTTCGATCCACCGCACAGCGCTGTGCTCTTCT 372
QY 111 CGAGAGGCGCGTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTGACACCTTCTGA 170
|||||
Db 373 CGAGAGGCGCGTGTGAGTCTCCAGTATGTTCCGCCAGCGCTCTTGACACCTTCTGA 432
QY 171 CCAAGCGCCAGGACTGCGAGCCAGGAGAGAGGGGCTCACCTTTATCTCTGGGAGCCC 230
|||||
Db 433 CCAAGCGCCAGGACTGCGAGCCAGGAGAGAGGGGCTCACCTTTATCTCTGGGAGCCC 492
QY 231 ACTGCACAGAGAGCGCGCTCTCCAGACTTAAATGTATCACCACCTAACCCTGTAGGGGG 290
|||||
Db 493 ACTGCACAGAGAGCGCGCTCTCCAGACTTAAATGTATCACCACCTAACCCTGTAGGGGG 552
QY 291 ACCCAATCTGGAATCTCTTCCCGCGCTTGGGACATCCGAGCGCGGGAAGCAGTGCCTCCCA 350
|||||
Db 553 ACCCAATCTGGAATCTCTTCCCGCGCTTGGGACATCCGAGCGCGGGAAGCAGTGCCTCCCA 612
QY 351 GGCCTGGGCCAGAGAGCTCCAGGAAGGCACTGAGCGCTGTGGCGGAGGCGCTCCGAC 410
|||||
Db 613 GGCCTGGGCCAGAGAGCTCCAGGAAGGCACTGAGCGCTGTGGCGGAGGCGCTCCGAC 672
QY 411 ATCCGAGGCAACAGGAAAGTCTCTGGGGCGGATCTGTAATAAACCCTTTTCTTTT 470
|||||
Db 673 ATCCGAGGCAACAGGAAAGTCTCTGGGGCGGATCTGTAATAAACCCTTTTCTTTT 732
QY 471 GTTTT 475
|||||
Db 733 TTTT 737

Search completed: February 23, 2004, 23:20:18
Job time : 1268.96 secs